THE UNIVERSITY
of EDINBURGH

The Cellular Functions of the Microprocessor Complex

Ross Andrew Alex Cordiner

Submitted for Doctor of Philosophy

The University of Edinburgh

September 2015

Declaration

This dissertation is the result of my own work unless otherwise stated. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma or other qualification.

Ross A. A. Cordiner

11th September, 2015

Acknowledgements

I want to thank a number of people who have helped me throughout my Ph.D., and also mention some of the highlights of the last four years which have been incredible.

When I initially approached Javier to enquire about a rotation project within his laboratory, I questioned why he was not currently supervising any students, to which he replied, “I hate students.” Undeterred by this response and always relishing a challenge I decided to peruse my Ph.D in his laboratory. Working in the Cáceres laboratory has provided me with an inspirational scientific environment, which has enabled me to work with many talented researchers. In addition, I have been allowed a great deal of academic freedom, with many opportunities to enhance my future career prospects, such as presenting at numerous international conferences. These opportunities may not have been available working within other groups; for this, I am grateful to Javier.

I’d like to especially thank Dr. Sara Macías, my supervisor throughout my Ph.D. Sara is a truly awe inspiring scientist, and her unwavering support (…and also *de gratis* beers) got me through all of the failed experiments and confidence crises. In truth without her guidance I would have not achieved half as much as I have in the last four years, and for that I will be forever grateful.

I would also like to thank past and present members of the Cáceres laboratory for creating vibrant and excellent academic environment to work in. A particular incident that is unrelated to science but that I’d like to mention would be the “3 ½ inch floppy disk” saga, which could have only occurred in a “Spanglish” speaking lab like ours. Special thanks go to Elena, Noemi, Nele, and Magda for keeping me entertained by giving me the nickname of “Rossini” and “hipster”, and being my drinking buddies at international conferences’.

Finally, I’d also like to thank my parents, Ross and Isabel, and my sisters, Kirsty and Lynsey, who have listened to all my problems and moaning on a weekly basis and have been incredibly understanding of my non-attendance to various family events; you’re the best!

"haters gonna, hate, hate, hate, hate, hate…"

Swift, Taylor (2014)

Abstract

DGCR8 (DiGeorge critical region 8) protein constitutes part of the Microprocessor complex together with Drosha, and is involved in the nuclear phase of microRNA (miRNA) biogenesis. DGCR8 recognises the hairpin RNA substrates of precursor miRNAs through two double-stranded RNA (dsRNA) binding motifs and acts as a molecular anchor to direct Drosha cleavage at the base of the pri-miRNA hairpin.

Recent characterisation of the RNA targets of the Microprocessor by HITS-CLIP of DGCR8 protein revealed that this complex also binds and regulates the stability of several types of transcripts, including mRNAs, lncRNAs and retrotransposons. Of particular interest is the binding of DGCR8 to mature small nucleolar RNA (snoRNA) transcripts, since the stability of these transcripts is dependent on DGCR8, but independent of Drosha. This raises the interesting possibility that there could be alternative DGCR8 complex/es using different nucleases to process a variety of cellular RNAs.

We performed mass spectrometry experiments and revealed that DGCR8 co-purifies with subunits of the nucleolar exosome, which contains the exonuclease RRP6. We demonstrated DGCR8 and the exosome form a nucleolar complex, which degrade the mature snoRNAs tested within this study. Interestingly, we also show that DGCR8/exosome complex controls the stability of the human telomerase RNA component (hTR/TERC), and absence of DGCR8 creates a concomitant telomere phenotype.

In order to identify the RNA targets of the DGCR8/Exosome complex on a global scale we performed iCLIP of endogenous and overexpressed RRP6 (wild-type and a catalytically inactive form). Thus, intersection of CLIP datasets from DGCR8 and RRP6 identified common substrates; accordingly snoRNAs were the most represented. In addition, we identified the cellular RNA targets of the RRP6 associated human exosome. The use of a catalytically inactive form of RRP6 stabilised important in vivo interactions that are highly dynamic and transient and also highlighted the role of RRP6-mediated

trimming of 3'flanks of immature non-coding RNAs. We will present a global view of the RNA-binding capacity of the RRP6-associated exosome.

In sum, we identified a novel function for DGCR8, acting as an adaptor to recruit the exosome to structured RNAs and induce their degradation. Moreover, we have identified DGCR8-depenedent substrates of the exosome and have demonstrated the requirement of RRP6 for 3' processing of ncRNAs.

Table of Contents

Table of Figures

Table of Tables

Chapter 1. Introduction

1.1 Part 1: The functions of the Microprocessor complex

1.1.1 MicroRNAs

MicroRNAs (miRNAs) are a large family of small RNAs between 21-23 nucleotides (nts) in length, which function as post transcriptional inhibitors of gene expression. They were initially discovered in *C.elegans* (Lee et al., 1993), and since then have also been found in both plants and metazoans, and their importance is exemplified by conservation of *let-7* (Lagos-Quintana et al., 2001; Lau et al., 2001; Lim et al., 2003a, 2003b; Pasquinelli et al., 2003), amongst others, from *C.elegans* through to humans, indicating an ancient metazoan origin. Advances in sequencing technologies and computing power have led to current estimates of the human miRNOME comprising ~2500 miRNAs (Griffiths-Jones, 2004), with a large repertoire of potential miRNAs that remain to be fully characterised (Friedländer et al., 2014).

1.1.1.1 miRNA function

Following a regulated biogenesis process, mature miRNAs are bound by a member of the Argonaute (AGO) protein family, which is part of a large ribonucleoprotein (RNP) complex called the RNA induced silencing complex (RISC). RISC-associated miRNAs act as a guide to target specific mRNAs for post-transcriptional translational repression. The specific targeting of RISC is conducted by complementarity of position 2-7 (relative to 5' end of the miRNA) to target mRNAs; this sequence is also known as the seed sequence (Fig. 1 A). Positions outside this sequence also contribute to target specificity, however, since perfect base pairing is not essential one miRNA can target multiple mRNAs. There are 4 different AGO proteins in humans, which facilitate translational repression by recruiting mRNA decay factors. Interestingly however, AGO2 can cleave mRNAs via slicer domain. Cleavage is promoted by high sequence complementarity between the miRNA and its cognate mRNA.

1.1.1.2 Canonical miRNA biogenesis

The majority of human miRNAs (~56%) are embedded within intronic sequences of protein coding genes, while ~23% are intergenic and 10% located in exonic sequences (Ghorai and Ghosh, 2014). These nascent transcripts are termed primary-

miRNAs (pri-miRNA). In all instances, pri-miRNAs are transcribed by RNA polymerase II (RNAPII) and fold to create stem-loop structures. The stems harbour the mature miRNA sequences, which are partially complementary and can create bulges within the stem due to incomplete base pairing. Furthermore, depending on their location relative to the terminal loop, miRNAs are termed as 5p and 3p, respectively (Fig. 1 B).

Canonical miRNA biogenesis proceeds via consecutive nuclear and cytoplasmic processing steps (Fig. 2). First, the nuclear step of miRNA biogenesis is catalysed by the Microprocessor complex, which cleaves the pri-miRNA at its base, releasing the hairpin RNA, which is called precursor miRNA (pre-miRNA)(Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). The minimally catalytically active Microprocessor complex is comprised of the double-stranded (ds) RNA binding protein, DGCR8 (DiGeorge syndrome Critical Region 8; also known as *pasha* in invertebrates), and the RNase III enzyme, Drosha (formerly known as RNASEN). The pre-miRNA products of Drosha processing have a 2 nt 3' overhang, characteristic of the RNaseIII cleavage. This overhang is recognised and bound by Exportin-5 (Exp5), which exports the pre-miRNA to the cytoplasm (Okada et al., 2009; Yi et al., 2003; Zeng and Cullen, 2004). Thereafter, another RNase III enzyme, Dicer, cleaves 22nt up from the

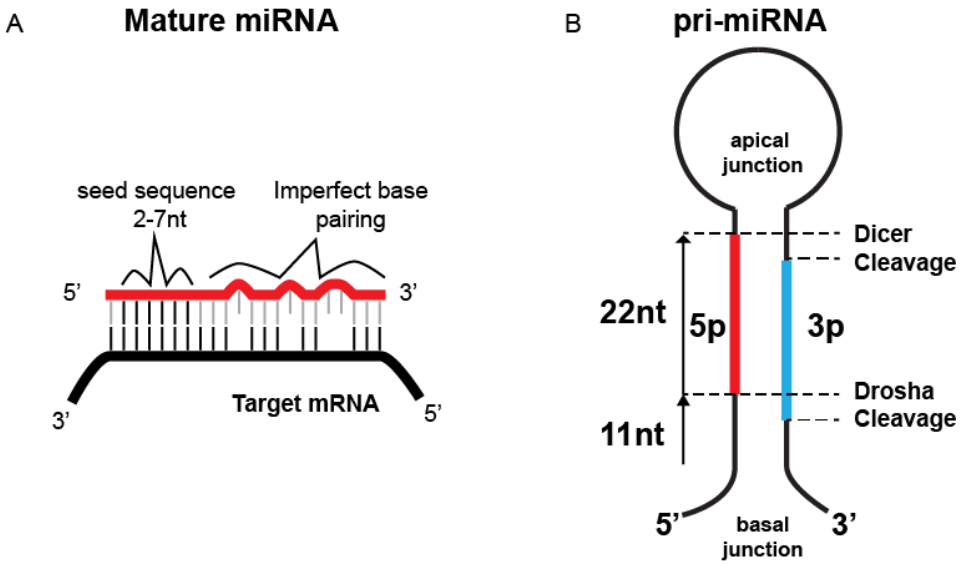


Figure 1 | Important features and structure of a miRNA. (A) The mature miRNA is 21-23nt long and the seed sequence is located from nucleotides 2 – 7, whereas other nucleotides can imperfectly base pair with the target mRNA. (B) The pri-miRNA is ~70nt long. Drosha cleavage is measured 11bp up from the ss-dsRNA basal junction, whereas Dicer processing is measured 22nt up from the 5' end created by Drosha processing.

5' end of the pre-miRNA (Park et al., 2011) and removes the loop structure leaving a ~22nt miRNA duplex. Dicer cleavage is facilitated by two closely related proteins, TRBP (TAR-RNA-binding protein) and PACT (protein Activator of the interferon- induced protein kinase). Although Dicer interacts with these two proteins, they are not required for

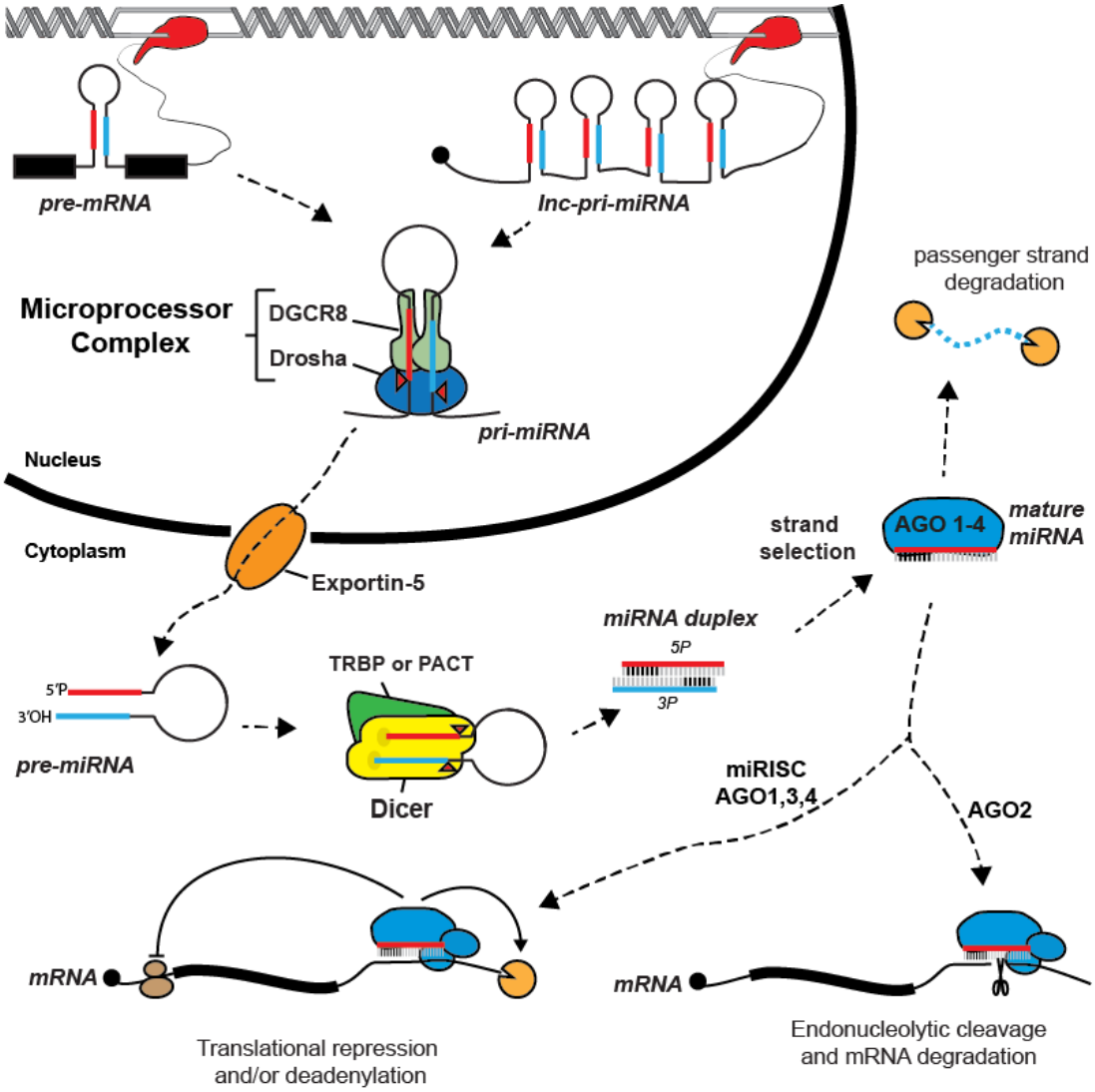


Figure 2 | The canonical miRNA biogenesis pathway. Primary miRNAs are contained predominantly within introns of host pre-mRNAs. In addition, pri-miRNA can be transcribed as an independent transcription unit, usually as part of a cluster of miRNAs (lnc-pri-miRNA). Canonical pri-miRNAs are transcribed by RNAPII (red protein on DNA), and are then processed by the Microprocessor complex. DGCR8 and Drosha form the minimally active Microprocessor complex. Drosha is the catalytic subunit which cleaves the pri-miRNA 11 bp up from ss-dsRNA junction, Drosha cleavage is denoted by red triangles on Drosha. The pre-miRNA is exported by Exportin-5 which recognises the 2nt 3' over hang created by imperfect Drosha cleavage. The pre-miRNA is bound by bound by Dicer and one of the cofactors, TRBP or PACT. Dicer binds 5' phosphate (5P) in PAZ pocket domain (dark yellow circle) and measures 22nt up from here. Dicer cleavage, denoted by red triangles, produces miRNA duplex containing 5P and 3P miRNA. Next, the miRNA duplex is bound by one of the Ago proteins and the mature strand is stably bound, whilst the other strand, termed the passenger or star strand, is ejected and degraded by cellular exonucleases. The AGO proteins, as part of the miRISC (miRNA RISC), use the mature miRNA as an antisense guide to target the miRISC to cognate mRNAs, which induces translational repression and or recruit deadenylation factors. Alternatively, AGO2 can endonucleolytically cleave and destabilise mRNAs via its slicer domain.

its activity, but have been implicated in defining the cleavage site and the formation of the RISC (Fukunaga et al., 2012; Lee et al., 2006). Next, the miRNA duplex is loaded into the Argonaute proteins (Ago) to generate the RISC complex (RNA Induced Silencing Complex; also known as miRISC [miRNA-RISC]). The Ago proteins then select one strand, which is termed the mature miRNA, whereas the other strand is discarded and degraded (termed the star/passenger strand)(Kobayashi and Tomari, 2015; Yao et al., 2015). The mature miRNA is then used as a guide, to target the RISC to mRNAs for translational repression and mRNA degradation (Fig. 2)(for review see Filipowicz et al., 2008; Kim et al., 2009).

1.1.1.3 Non-canonical miRNA biogenesis

Canonical miRNA biogenesis is driven by the RNase III enzymes, Drosha and Dicer. However, it is now established that other non-canonical biogenesis pathways can produce functional miRNAs by means of mechanisms other than RNase III cleavage. Therefore, this has resulted in a broadening of miRNAs classification. A factor common to all non-canonical miRNAs identified to date, is they require cleavage by at least one RNaseIII enzyme involved in canonical miRNA biogenesis. Thus, non-canonical miRNAs can be subdivided into their processing dependencies: Dicer-dependent or Microprocessor-dependent (summarised in Fig. 3) (for review see Yang and Lai, 2011).

1.1.1.3.1 Dicer-dependent non-canonical miRNA biogenesis

1.1.1.3.1.1 Mirtrons

One such class of non-canonical miRNAs are mirtrons, which are embedded within intronic sequences of host pre-mRNAs (Fig.3 A). In contrast to canonical miRNAs, host-intronic sequences of mirtrons are short and their initial step of biogenesis is mediated by pre-mRNA splicing (Okamura et al., 2007; Ruby et al., 2007). Specifically, conventional mirtrons 5' and 3' precursor sequences coincide with splice donor and acceptor sequences, and as a consequence, splicing and subsequent debranching of the lariat facilitates their release, enabling the mirtron to fold into a canonical pre-miRNA structure. Thereafter, the mirtron is transported to the cytoplasm and follows the canonical miRNA biogenesis pathway by Dicer processing. In addition to conventional mirtrons, tailed mirtrons have also been identified (Babiarz et al., 2011; Okamura et al., 2007). Their

biogenesis is still splicing dependent; however, they differ from conventional mirtrons as their precursor sequence 5' or 3' termini are defined by exonucleases in order to create a viable pre-miRNA structure for Dicer processing (Flynt et al., 2010).

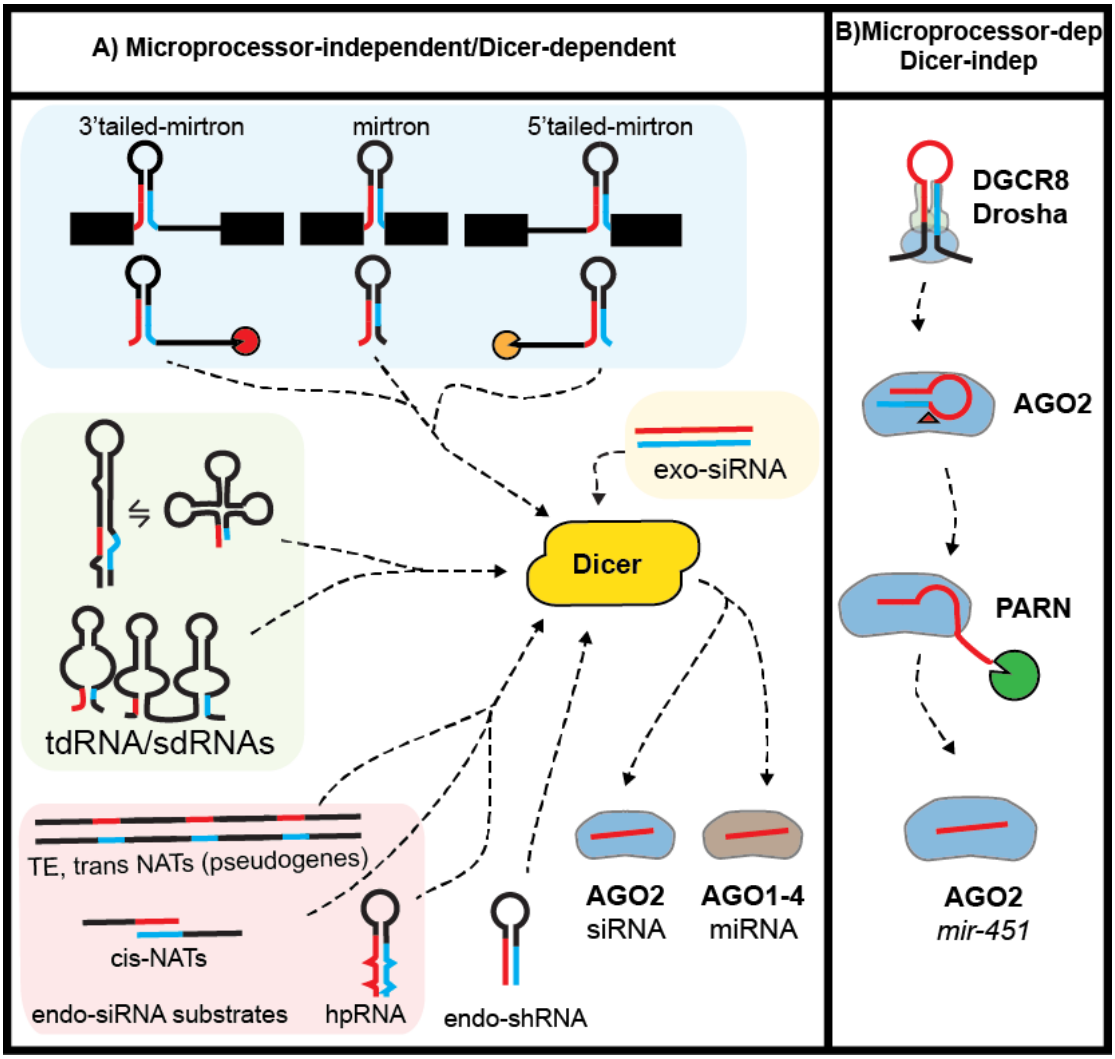


Figure 3 | Non-canonical miRNA biogenesis & siRNA biogenesis. (A)Most non-canonical miRNA/siRNAs biogenesis is Dicer-dependent. The release of mirtrons from intronic sequences is mediated by splicing and debranching of the resulting lariat, thereafter the pre-mirtron which is then processed by Dicer. Tailed mirtrons 3' or 5' ends are defined by exonucleases after splicing, in order to generate a viable Dicer substrate. Some tRNAs and snoRNAs have alternative conformations which can be processed by Dicer into sdRNAs or tdRNAs, respectively, which can regulate alternative splicing or gene expression via RNAi. Additionally, endo-siRNAs are diced from trans or cis natural antisense transcripts (NATs) which are produced by convergent transcription. Hairpin (hp) RNAs are produced from long cis transcribed 3'UTRs which can refold to form large hairpin structures, whereas endo-shRNAs are transcribed by RNAPIII from their own promoters and at least one of their termini is determined by transcription start or stop sites. Exo-siRNAs (exogenous-siRNAs) are produced from exogenous viral transcripts. After Dicer processing, the miRNAs are loaded into one of the four AGO proteins, whereas siRNAs are specifically loaded into AGO2. (B) The pre-miRNA from mir-451 is loaded directly into AGO2 which cleaves below the terminal loop on the 3p arm and then the exonuclease PARN, trims away the remaining non-miRNA sequence. This figure was adapted from Yang & Lai, (2011).

1.1.1.3.1.2 Small RNAs derived from snoRNAs and tRNAs

Sequencing studies of Ago protein-bound RNAs aimed at identifying putative miRNAs have also uncovered a number of small RNAs derived from tRNAs (tdRNAs) and snoRNAs (sdRNAs) (Fig.3 A). One such example of snoRNA-derived RNAs is *ACA45* snoRNA/*mir-1839*, which was subsequently found to be Dicer-dependent but DGCR8/Drosha-independent (Ender et al., 2008). Deep sequencing of *Dgcr8* or *Dicer1* deficient mESC identified heterogeneous patterns of reads the around the *tRNA-Ille/mir-1983* locus (Babiarz et al., 2008). The mir-1983 present in the 3' end of the tRNA was found to be DGCR8-independent, but Dicer-dependent. Furthermore, the precursor of tRNA-Ile/mir-1983 was predicted to fold into an alternative confirmation creating a viable Dicer substrate. Therefore, different structural conformations may dictate entry into alternative biogenesis pathways. Further deep-sequencing studies aimed at addressing the frequency and abundance of small RNAs of derived ncRNAs identified a large population of highly abundant RNA fragments, some of which were sensitive to DGCR8 or Dicer deficiencies (Babiarz et al., 2008, 2011; Cole et al., 2009; Haussecker et al., 2010; Taft et al., 2009). However, as Dicer processing does not automatically grant Ago loading, and these fragments may represent stable degradation intermediates; further functional studies are needed to address the contribution of sd/tdRNAs to RNA silencing.

1.1.1.3.2 Microprocessor-dependent non-canonical miRNA biogenesis

The only Microprocessor-dependent, Dicer-independent miRNA is the erythropoietic *mir-451*(Fig.3 B)(Cheloufi et al., 2010). This miRNA is processed by the Microprocessor, generating a pre-miRNA too short for Dicer processing. Alternatively, the pre-miRNA is loaded directly into Ago2, which slices the pre-miRNA generating a 30nt intermediate which is further processed by the 3'→5' exonuclease PARN (PolyA-specific RiboNuclease)(Yoda et al., 2013).

1.1.1.4 Other mechanism of RNAi generated by the miRNA biogenesis machinery

The metazoan small RNA family is comprised of miRNAs, siRNAs (small interfering RNAs) and piRNAs (piwiRNAs) (Table 1)(for review see Ghildiyal and Zamore, 2009). The functions of Dicer are not restricted to miRNA biogenesis as the catalytic action is

also required for siRNA biogenesis. A series of dsRNA such as, endogenous siRNAs (endo-siRNAs), exogenous siRNAs (exo-siRNA), short hairpin siRNAs (shRNAs) and natural antisense transcripts (NATs) have all been identified to be Dicer-dependent and can function as antisense guides for the RISC machinery (Fig. 3)(Babiarz et al., 2008, 2011).

1.1.2 The Microprocessor complex

The Microprocessor complex is fundamental for the biogenesis of canonical miRNAs in higher eukaryotes, exemplified by its conservation together with miRNAs from humans to worms (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). However, miRNAs and Microprocessor components are absent from *S.cerevisiae*. As such, it seems the Microprocessor complex and miRNAs co-evolved after the divergence of our last common ancestor with budding yeast. Nevertheless, a single RNase III protein present in budding yeast; Rnt1p shares some common functions with the Microprocessor (Dhir et al., 2015; Rondón et al., 2009), although the domain structure is far more basic, consisting of only one dsRBD and a single RNase III domain, (MacRae and Doudna, 2007).

Table 1 | Mammalian RNAi classes. Table originally published in Ha and Kim et al. (2014)

Features	**miRNA**	**siRNA**	**piRNA**
Length	~ 22 nt	~21 nt	24-30nts
Processing enzymes	Microprocessor and Drosha	Dicer	Zucchini and unknown trimming enzymes
AGO subclades	AGO 1-4	AGO 2	PIWI
Mechnaism of action	Translational repression mRNA degradation	RNA cleavage	TGS or PTGS of transposons Multigenerational epigenetic phenomena in worms
Function	Regulation of protein coding genes	Regulation of protein-coding genes Regulation of transposons Antiviral defence	Pre-pachytene piRNA: transposon silencing Pacytene piRNA: unknown piRNA-like small RNA in soma :unknown

1.1.2.1 The Microprocessor complex exists as two different molecular weight complexes

The initial characterisation of the Microprocessor complex identified two different molecular weight complexes; a smaller 440kDa, containing DGCR8 and Drosha that was subsequently shown to be necessary and sufficient for pri-miRNA processing; and a larger >2MDa complex, which contained DGCR8 and Drosha, plus 20 additional proteins identified by mass spectrometry (Gregory et al., 2004). These associated proteins include RNA helicases, heterogeneous ribonucleoproteins (hnRNPs) and proteins with RNA-binding activity, which have subsequently been shown to modulate Microprocessor activity. Specifically, depletion of p72 and p62, two RNA helicases identified in the mass spectrometry analyses, caused a concomitant depletion in mature miRNAs suggesting that these additional Microprocessor associated factors can promote miRNA biogenesis (Gregory et al., 2004).

1.1.2.2 Essential components of the Microprocessor

The smaller 440kDa Microprocessor complex contains DGCR8 and Drosha (Gregory et al., 2004; Nguyen et al., 2015). As these components are essential for *in vitro* and *in vivo* activity of the Microprocessor this is termed the "minimal" Microprocessor complex. The interaction between DGCR8 and Drosha is mediated through the C terminal tail of DGCR8, which is also required for the stability of Drosha (Fig. 4 A)(Han et al., 2004, 2009; Nguyen et al., 2015; Yeom et al., 2006). Interestingly, DGCR8 also contains a RNA-binding heme-domain (Rhed), which is necessary for miRNA biogenesis (Faller et al., 2007). The Rhed binds to pri-miRNAs and may be required for DGCR8 homodimerisation (Nguyen et al., 2015; Quick-Cleveland et al., 2014; Weitz et al., 2014). The RNA-binding activity of DGCR8 is provided by two adjacent dsRNA binding domains (dsRBD). Drosha also possess a dsRBD, which is necessary but not sufficient for correct miRNA processing (Lee et al., 2003). Therefore, DGCR8 is considered to be the RNA binding component of the Microprocessor complex, whereas Drosha is the catalytic component. Drosha is a group two RNase III protein and contains two RNase III domains, a dsRBD and a P-rich and RS-rich domain (Fig. 4 B). The region of Drosha required for the interaction with DGCR8 is located between the RS-rich and RNase III domains (Han

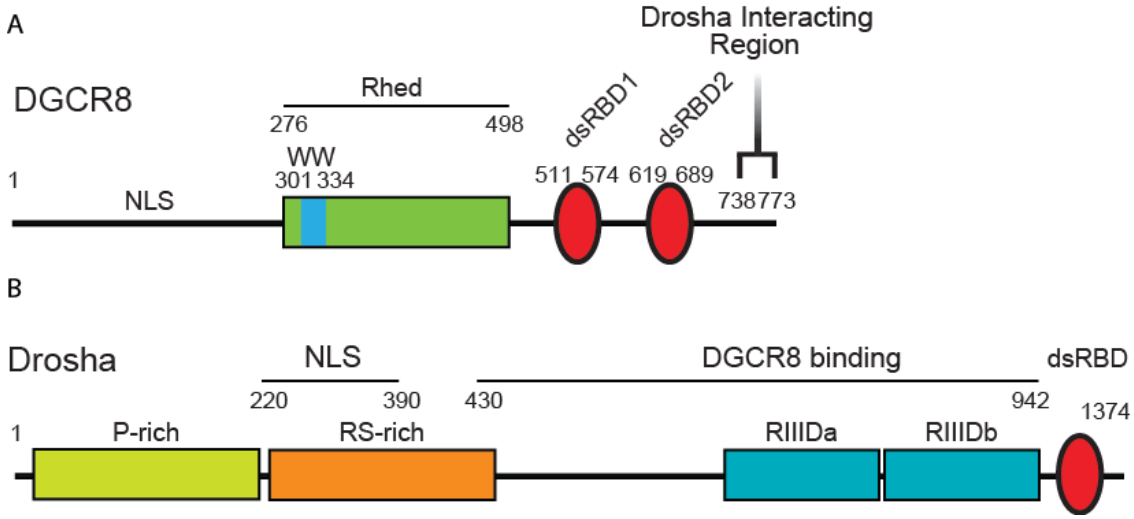


Figure 4 | DGCR8 and Drosha Schematics. DGCR8 is 773 amino acids long and contains an nuclear localisation signal (NLS) in it N terminal domain; an RNA heme binding domain (Rhed; green) and a proline rich domain (WW; blue) which has been shown to be important for miRNA processing and homodimerisation; two adjacent dsRNA binding domains (dsRBD1 and dsRBD 2; red); and, a C-terminal tail region which is important for Drosha interaction (738-773; Drosha interacting region), and is required for Drosha stability. Drosha contains; a proline-rich (P-rich; green) and adjacent arginine and serine-rich (RS-rich; orange) domains, which are important protein:protein interactions; an RNaseIII domain, which forms a intramolecular dimer around dsRNA substrates and offsets its two active sites (RIIIDa and RIIIDb), to produce 2nt 3' overhangs with 5' phosphates and 3' OH ends, typical of RNase III enzymes; and a C terminal dsRBD. The NLS of Drosha spans 220-390, whereas the region required for DGCR8 binding spans amino acids 430-942.

et al., 2004; Nguyen et al., 2015). The RNase III domains form an intramolecular dimer around dsRNA substrates and the catalytic sites are imperfectly aligned, thus resulting in a cleavage that creates a 2nt 3' overhang, which is characteristic of group two RNase III proteins (Han et al., 2004; MacRae and Doudna, 2007). This imperfect cleavage by Drosha is important since the 2nt 3' overhang is recognised by both Exp5 and Dicer in subsequent steps of miRNA biogenesis (Okada et al., 2009; Park et al., 2011). Furthermore, Drosha processing defines the 5' and 3' ends of the pre-miRNA, which is important on two levels. Firstly, Dicer binds the phosphate of 5' termini of the pre-miRNA and cleavage is directed 22nts away, therefore Drosha predefines the site cleaved by Dicer (Fig. 1 B)(Park et al., 2011). Secondly, as Drosha processing directs Dicer cleavage, it also controls the sequence of the resulting miRNA. Given the importance of complementarity between the miRNA seed sequence and cognate mRNA target, shifts of 1 nt within the seed sequence can change the repertoire of cognate mRNA targets (Helwak et al., 2013). Therefore understanding the mechanism that directs Drosha cleavage is of extreme importance.

1.1.2.3 The substrate recognition mechanism by the Microprocessor complex

Despite intensive analysis of miRNAs function and the discovery of alternative biogenesis pathways, the fundamental mechanism of DGCR8 binding to its RNA substrates remains controversial. Originally, it was reported that DGCR8 recognises the basal ss-dsRNA junction (single-stranded-double-stranded RNA)(Fig. 1 B), anchoring the Microprocessor to its substrate, and then directs Drosha approximately 1 helical turn (11bp) up to cleave (Han et al., 2006). In addition, the number of DGCR8 molecules constituting the Microprocessor complex was poorly understood until recently. Two models of binding were proposed: firstly, DGCR8 binds solely to the apical ss-dsRNA junction as a homodimer (Nguyen et al., 2015); alternatively, DGCR8 can bind to both apical and basal ss-dsRNA junctions of the pri-miRNA and this involves two homodimers of DGCR8 (Quick-Cleveland et al., 2014).

The first hypothesis was drawn from observation of *in vitro* processing assays which demonstrated that Drosha does not require the apical loop to cleave a pri-miRNA substrate (Nguyen et al., 2015). In addition, analyses of pri-miRNA substrates which had their UG and UGU motifs swapped with each other or deleted, showed that they are required for correct Microprocessor/substrate orientation. This indicates that Drosha interacts with the basal UG motif whereas DGCR8 interacts with the apical UGU motifs (Nguyen et al., 2015). These motifs are embedded within pri-miRNA and represent sequence elements that were found to promote pri-miRNA processing (Auyeung et al., 2013). Contrarily, this suggests that DGCR8 recognises sequence rather than structural elements. However, 55% of all known human miRNAs lack either of these sequences, and only 7% contain both motifs, arguing against these motifs being a general requirement for all miRNAs to be processed by the Microprocessor.

The second proposed binding mechanism is compiled from the observations that DGCR8 mutants deficient in heme binding display pri-miRNA processing defects (Faller et al., 2007; Quick-Cleveland et al., 2014). The Rhed domain within DGCR8 binds an essential heme (protoporphyrin IX in complex with iron) co-factor and is required for homodimerisation (Faller et al., 2007). Size exclusion chromatography assays

demonstrated that those conformations of pri-miRNA-DGCR8 complexes constituted from DGCR8 heme-binding mutants were different to wild-type DGCR8 complexes. In addition, *in vitro* binding assays identified that pri-miRNAs contain two putative DGCR8 binding sites, at the basal and apical ss-dsRNA junction. The authors concluded that heme binding by DGCR8 is required for clamping of DGCR8 homodimers to the pri-miRNA. However, as these experiments identifying the requirement for heme and miRNA processing were identified *in vitro,* open questions remain over the requirement for heme within miRNA biogenesis *in vivo,* and whether heme availability in cells can regulate the Microprocessors function.

1.1.2.4 Regulation of the Microprocessor complex

MiRNA expression is regulated at every level, from transcription to biogenesis and turnover. Dynamic regulation is required to achieve specific temporal miRNA expression to regulate an intricate network of gene expression. Specifically, within the nuclear phase of miRNA biogenesis, the Microprocessor processing step is highly regulated (for review see Finnegan and Pasquinelli, 2013; Ha and Kim, 2014).

The components of the Microprocessor complex, DGCR8 and Drosha, post-transcriptionally regulate each other. DGCR8 protein is required to stabilise Drosha at the protein level, and, DGCR8 levels are regulated by the Microprocessor at the mRNA level (Han et al., 2009; Kadener et al., 2009; Nguyen et al., 2015). The Microprocessor complex binds and cleaves two pri-miRNA-like hairpins embedded within *Dgcr8* mRNA. Accordingly, depletion of Drosha results in a ~2 fold increase in *Dgcr8* mRNA, whereas depletion of DGCR8 protein results concomitant decrease of Drosha protein levels. This mechanism of auto-regulatory feedback ensures homeostatic levels of the Microprocessor are maintained.

Furthermore, the individual components of the Microprocessor complex can also undergo post-translational modifications which alters miRNA biogenesis (summarised in Fig. 5). For example, phosphorylation by GSK3β is required for nuclear localisation of Drosha (Tang et al., 2010, 2011); whereas, acetylation (Tang et al., 2013) or association of TDP43 protein stabilises Drosha (Di Carlo et al., 2013; Kawahara and Mieda-Sato, 2012). DGCR8 can also be SUMOylated (Zhu et al., 2015) and

phosphorylated which increase protein stability and also promote the production of pro-growth miRNAs (Herbert et al., 2013). In addition, acetylation of DGCR8 by HDAC1 increases its affinity for pri-miRNAs (Wada et al., 2012). DGCR8 activity can also be modulated by MECP2 and ADAR, which sequester DGCR8 protein, therefore preventing proper Microprocessor complex function (Cheng et al., 2014; Nemlich et al., 2013).

1.1.3 Genetic studies of miRNA biogenesis factors reveals non-canonical functions of the Microprocessor complex

Phenotypic comparisons of cell lines and tissues deficient for the miRNA biogenesis machinery components *dgcr8, drosha, dicer* and *ago* were expected to yield overlapping phenotypes, if a common pathway is affected (miRNA biogenesis). However, given the abundance of Dicer-dependent, DGCR8/Drosha-independent RNAs that participate in RNAi (e.g., endo-siRNAs, mirtrons), the loss of Dicer should give rise to different phenotypes when compared with *dgcr8* or *drosha* mutants. Likewise, RNA substrates regulated exclusively by DGCR8/Drosha or Dicer should yield diverse

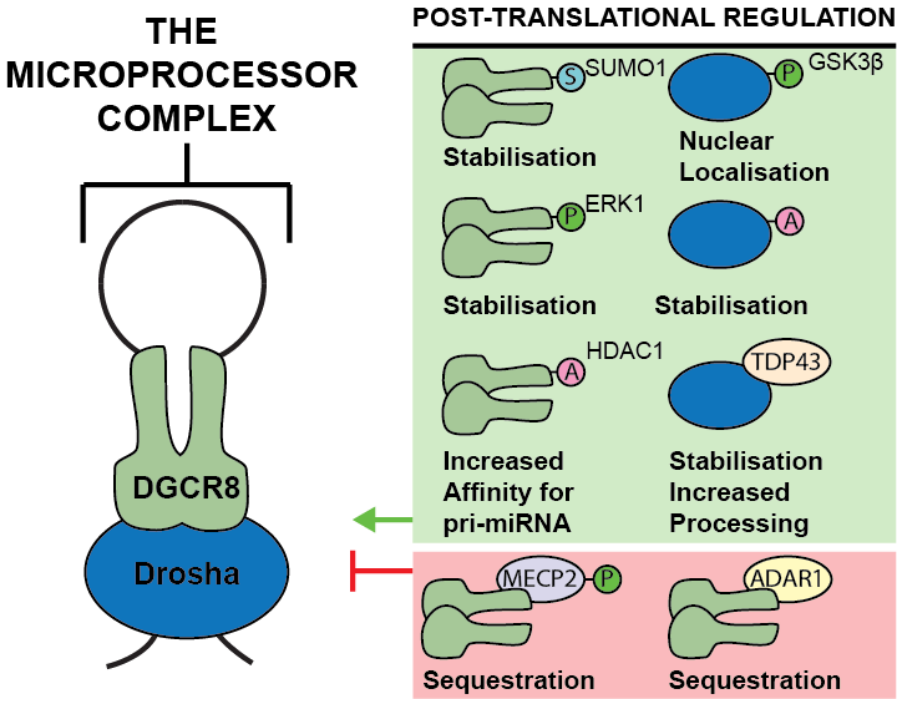


Figure 5 | Post-translational regulation of the Microprocessor. Schematic of the minimally active Microprocessor complex bound to pre-miRNA. Protein factors that promote (green box) or inhibit (red box) the Microprocessor activity and function of the post-translational regulation shown in bold below. Proteins that post-translationally modulate DGCR8 and Drosha activity are shown as interacting or modifying DGCR8 or Drosha. P, Phosphorylated; A, acetylated; S, SUMOylated.

phenotypes. When the phenotypes of the genetic mutants were compared, in some cases, deficiencies of the miRNA biogenesis factors gave rise to redundant phenotypes, supporting the notion that the common affect pathway is miRNA biogenesis. Nevertheless further genetic studies have uncovered non-redundant phenotypes with differing degrees of severity, highlighting important miRNA-independent roles for *dgcr8*, *drosha* and *dicer* proteins within a range of different cell and tissue types. Several examples of these studies are discussed below.

1.1.3.1 miRNA-dependent functions of DGCR8, Drosha and Dicer

Functional studies of the miRNA biogenesis machinery using genetic knockout mouse models established that *dgcr8*, *drosha* and *dicer* deficiencies all result in early embryonic lethality, confirming a critical role for miRNA in normal development (Bernstein et al., 2003; Murchison et al., 2005; Wang et al., 2007). Next, these studies were extended to differentiated cells, where phenotypic comparisons of natural killer cells (Bezman et al., 2010), regulatory thymocyte cells (Treg)(Chong et al., 2008), epidermal and hair follicle epithelial cells (Yi et al., 2009), which were deficient in *dgcr8/drosha* and *dicer* also presented similar phenotypes. These observations logically support that these phenotypes are produced by a common affected pathway, the canonical miRNA biogenesis, which demonstrates the importance of miRNA-mediated gene regulation in a number of cellular contexts.

1.1.3.2 miRNA-independent RNAi-dependent function of Dicer

Studies with *Dicer -/-* and *Dgcr8 -/-* mESC identified miRNAs as critical regulators for silencing mESC cell markers during differentiation, as both cell lines failed to differentiate *in vitro* and retained expression of pluripotency markers (Kanellopoulou et al., 2005; Wang et al., 2007). However, the *Dicer* deficiency phenotype displayed a more resolute block in differentiation and slower growth rates. Moreover transcriptomic comparisons of *Dgcr8 -/-* and *Dicer -/-* identified a multitude of Dicer-dependent, Microprocessor-independent sRNAs, such as mirtrons, endo-siRNAs and shRNAs (small hairpin RNAs)(Babiarz et al., 2008). This could explain the more severe phenotype observed in the *Dicer-/-* mESC. Interestingly, these analyses were also extended to post-mitotic neuronal cells and oocytes which recapitulated the more severe Dicer phenotype

(Babiarz et al., 2011; Suh et al., 2010). In sum, as Dicer processing is essential for the majority of RNAi/miRNA biogenesis pathways, this explains the severity of the *Dicer*-null phenotype in the cell lines and tissues outlined above.

1.1.3.3 RNAi-independent functions of Dicer

Recently, non-canonical functions of Dicer were revealed in a study of geographic atrophy (GA), an age related macular degeneration disease of the retinal pigmented epithelium (RPE). The authors identified reduced Dicer protein and mRNA levels, which led the authors to systematically analyse conditional knockouts of miRNA biogenesis pathway members within RPE cells (Kaneko et al., 2011). Intriguingly, only *Dicer* depletion recapitulated the GA phenotype. They subsequently showed that Alu repeat RNAs accumulate in the eye of GA patients' and Dicer serves to directly destabilise these transcripts. These findings were corroborated recently, by the identification that Alu RNAs are a major substrate of Dicer within humans and worms (Rybak-wolf et al., 2014). Interestingly, these CLIP experiments also revealed that Dicer could bind and stabilise a number of mRNAs, which suggest Dicer may also have important functions independent of its catalytic activity.

1.1.3.4 miRNA-independent function of Drosha

Phenotypic comparisons of early-stage T cells deficient of both *drosha* and *dicer* displayed a resolute block in differentiation, which was a more severe phenotype than their respective single deficiencies (Chong et al., 2010). These non-redundant phenotypes suggested that additional pathways were being affected by the double deletion. Subsequent transcriptomic studies identified a number of mRNAs that harboured pri-miRNA-like hairpins which were regulated directly by Drosha, but not Dicer (Chong et al., 2010). Interestingly, these observations regarding non-redundant phenotypes were not observed in later stage T regulatory cells (Treg), suggesting that Drosha-mediated mRNA regulation is important at specific stages of T cell development (Chong et al., 2008). Taken together, these observations support a miRNA-independent roles for the Microprocessor component Drosha.

1.1.3.5 miRNA-independent function of DGCR8

A common theme between all the studies outlined above is the comparison of DGCR8- or Drosha- with Dicer- deficiencies. The assumption that deficiencies in either of the Microprocessor components leads to redundant phenotypes had never been addressed within parallel genetic studies until recently. Individual deficiencies of *pasha* and *drosha* in neuronal cell lineage produces abnormalities that were not seen in *argonaute-1* or *dicer* deficient *D.melanogaster* (Luhur et al., 2014). Moreover the phenotypes displayed by *drosha* and *pasha* were not redundant. The mechanisms underlying the different phenotypes have yet to be explored. In addition, another study uncovered Drosha-independent and miRNA-independent functions of DGCR8 (Macias et al., 2012), which will be expanded on later in an additional section (1.5).

Altogether, the studies outlined above have emphasised the importance of miRNA-independent functions of the miRNA biogenesis machinery. In addition they have highlighted a previously underappreciated role for the Microprocessor components, DGCR8 and Drosha, suggesting they may also constitute other complexes that are important for developmental functions in differentiated tissues.

1.1.4 Non-canonical functions of the Microprocessor

Genetic studies have evidenced the Microprocessor complex has non-canonical function which are important in certain cellular and developmental contexts. Additional studies addressing the molecular mechanisms underlying these non-canonical functions have identified the Microprocessor complex can regulate a range of cellular processes. Interestingly, regulation of these process do not always require the catalytic activity of the Microprocessor complex. Examples of non-canonical functions and their biological relevance are discussed within this subsection (summarised in Fig. 6).

1.1.4.1 mRNA destabilisation

The best characterised non-canonical target of the Microprocessor is the *Dgcr8* mRNA. Drosha cleaves two pri-miRNA-like hairpins embedded within *Dgcr8* mRNA, leading to its destabilisation (Han et al., 2009; Kadener et al., 2009). This cleavage therefore serves as a mechanism of direct negative regulation. Further studies have also

identified a number of mRNAs containing pri-miRNA-like structures which are cleaved by the Microprocessor (Chong et al., 2010; Heras et al., 2013; Knuckles et al., 2012; Macias et al., 2012; Seong et al., 2014). Comparative transcriptomic studies of *Drosha* and *Ago2* deficient mESC identified a number of Drosha-dependent, Ago-independent mRNA cleavage products that coincided with hairpin structures. These observations were also extended to flies and humans. Within *D.melanogaster* S2 cells, *drosha* and *dicer-1* knockdowns identified 25 additional mRNAs whose regulation was dependent on Drosha, but not Dicer (Kadener et al., 2009). Additionally a number of mRNAs containing pri-miRNA-like structures were also found to be regulated in a similar nature within early-stage T cells (Chong et al., 2010). However, the breadth of RNAs substrates of DGCR8 was poorly understood until the direct genome wide approach of HITS-CLIP (Cross-linking immunoprecipitation coupled with high throughput sequencing) was applied to DGCR8, which identified a large repertoire of mRNAs and other RNAs that ware bound by Microprocessor complex (Macias et al., 2012).

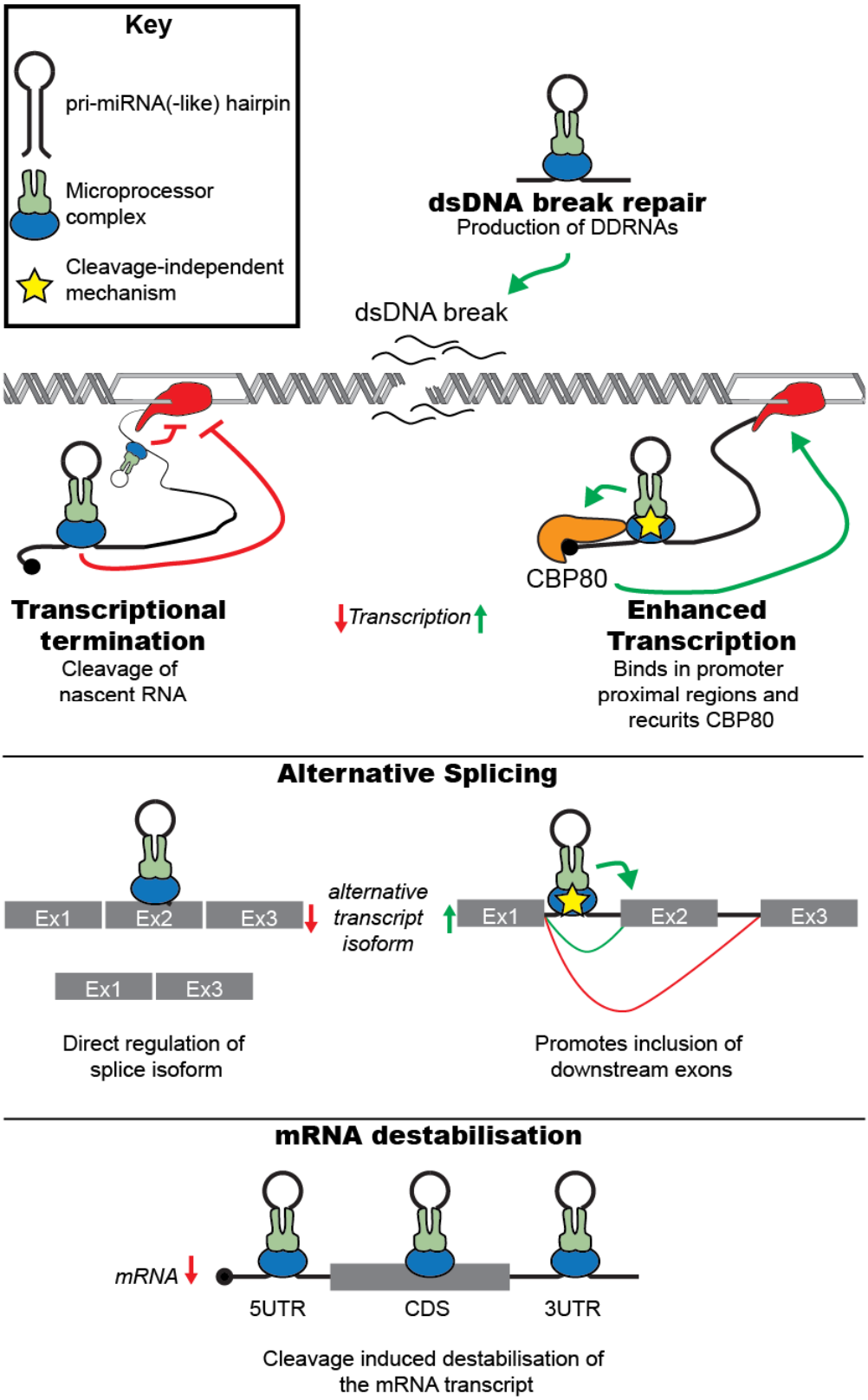
Direct regulation of mRNA transcripts by the Microprocessor was found to have a profound physiological significance during mouse neurogenesis, where conditional knockouts (c-KO) of *Drosha* and *Dicer* within neural stem cells (NSC) exhibited non-overlapping phenotypes (Knuckles et al., 2012). It was subsequently shown that c-KO of *Drosha* caused an upregulation of pro-differentiation neural mRNAs causing a concomitant upregulation of Ngn2 (Neurogenin 2) protein. Accumulation of this protein caused precocious differentiation of NSC; resulting in defective progenitor stem cell maintenance.

Interestingly, the canonical functions of the Microprocessor have been commandeered to regulate a viral mRNA from the Karposi's sarcoma-associated herpesvirus (KSHV)(Lin and Sullivan, 2011). KSHV can shift between a lytic and latent infection phase, which are characterised by low and high level KapB (Kaposin B) expression. The pre-mRNA of KapB contains a miRNA embedded within its 3'UTR and Drosha-mediated processing of this pri-miRNA causes concomitant mRNA destabilisation. Therefore the Microprocessor canonical functions regulate the viral life cycle KHSV.

Taken together, these studies support an essential role for direct regulation of mRNAs by the Microprocessor complex.

1.1.4.2 Transcriptional control

The Microprocessor was recently reported to facilitate enhanced transcription in a cleavage-independent nature (Gromak et al., 2013). It was reported that DGCR8 and Drosha can associate to nascent RNAs transcribed from promoter proximal regions of a number of human genes. From this location, Drosha promotes transcription and functions as a landing pad for the transcription initiation cofactor, CBP80 and the RNAPII (Fig. 6). In addition, two other studies involve the Microprocessor complex in transcription termination of HIV-encoded transcripts (Wagschal et al., 2012) and long ncRNAs containing pri-miRNAs (lnc-pri-miRNAs)(Fig. 6)(Dhir et al., 2015). The first involved RNAPII pausing and premature termination which is mediated by recruitment of the Microprocessor and the respective 5'→3' and 3'→5' exoribonuclease, Xrn2 and Rrp6 (Wagschal et al., 2012). Essentially, Microprocessor cleavage creates an internal entry point for both exoribonucleases to degrade the nascent transcript. An interesting finding from this study was that Rrp6 generates a functional sRNA that associates with the HIV-1 promoter to inhibit transcription. Therefore the Microprocessor complex can participate in both post-transcriptional gene silencing (PTGS) and also transcriptional gene silencing (TGS). The second study identified the Microprocessor complex is required for the transcriptional termination of lnc-pri-miRNA (Dhir et al., 2015). Unlike mRNAs, some lncRNAs do not contain the conventional polyadenylation site (PAS) motif, which promotes downstream cleavage and polyadenylation. Alternatively, for these linc-RNAs, Microprocessor-mediated cleavage of pri-miRNAs embedded within their sequence induces RNAPII termination and prevents transcriptional read-through into neighbouring genes (Dhir et al., 2015). Interestingly, a similar mechanism, was also found in *S.cerevisiae*, where the RNaseIII protein Rnt1p, can induce transcription termination (Rondón et al., 2009). Although the Microprocessor complex is not conserved in yeast, this suggests a conserved mechanism for RNaseIII-mediated transcriptional termination.


Key
pri-miRNA(-like) hairpin
Microprocessor complex
Cleavage-independent mechanism
dsDNA break repair
Production of DDRNAs
dsDNA break
CBP80
Transcriptional termination
Cleavage of nascent RNA
Transcription
Enhanced Transcription
Binds in promoter proximal regions and recurits CBP80
Alternative Splicing
Ex1
Ex2
Ex3
alternative transcript isoform
Ex1
Ex3
Ex1
Ex2
Ex3
Direct regulation of splice isoform
Promotes inclusion of downstream exons
mRNA destabilisation
mRNA
5UTR
CDS
3UTR
Cleavage induced destabilisation of the mRNA transcript

1.1.4.3 DNA Damage Response RNAs

The integrity of the genome is under constant threat by a myriad of chemical and environmental insults; causing dsDNA breaks. To rectify this, the DNA damage response apparatus is recruited to these sites to repair dsDNA breaks. Upon detection of DNA damage, γH2AX accumulate at the site of the dsDNA breaks, serving as a beacon to recruit DNA damage response (DDR) proteins. Interestingly, miRNAs and a class of Drosha- and Dicer- dependent sRNAs have been shown to be recruited to the site of DNA damage and participate in the repair process (Fig. 6)(Francia et al., 2012). These RNAs, termed DDRNAs (DNA Damage Response RNAs), were of the expected size after Drosha and Dicer products and their sequences matched the DNA legion sites. However, DDRNAs are distinct from canonical miRNAs; they show no complementarity to cellular mRNAs and depletion of components of RISC, which are essential for translational repression mediated by miRNAs, did not affect DDRNAs function. It is also interesting to note that increases in γH2AX foci seen in Drosha and Dicer Knockout hair follicle, could be explained by the loss of DDRNAs (Teta et al., 2012).

1.1.4.4 Microprocessor-mediated control of alternative splicing

The Microprocessor complex can also influence alternative splicing. Interestingly, opposing roles, both promoting and negatively regulating cassette exon inclusion have been reported by independent groups (Fig. 6). The identification of the direct RNA targets of DGCR8 by Macias et al. (2012) identified DGCR8 could bind a number pri-miRNA-like hairpins embedded within cassette exons of coding mRNAs. In this example, Microprocessor binding resulted in cleavage of the transcript isoforms containing the cassette exon, therefore changing the relative abundance between exon included/skipped transcript isoforms. Furthermore, it was recently reported the processing of pri-miRNAs encoded over intron-exon boundaries of cassette exons resulted in a

Figure 6 | The Non-canonical functions of the Microprocessor complex. Summary of the molecular mechanisms of the non-canonical functions of the Microprocessor complex (see bold for biological processes). The Micrprocessor complex (and also Dicer) are required for the production of sRNAs that associate with dsDNA breaks. In addition Microprocessor-cleavage is required for the transcriptional termination of lnc-pri-miRNA, and also the premature termination transcripts transcribed from the HIV-1 promoter. On the other hand, the Microprocessor complex has been shown to bind hairpins in nascent RNA and recruit CBP80, which in turn binds the 5' cap (black dot) and enhances transcription. Importantly however, this function of the Microprocessor does not require cleave (star). Similarly, the Microprocessor can promote inclusion of cassette exons by binding hairpins in the intronic sequences of pre-mRNAs. Again, this function of the Microprocessor complex is independent of its cleavage activity. The Microprocessor can also negatively regulate mRNAs that contain viable pri-miRNA-like hairpins throughout their sequence. This mechanism requires the endonucleolytic cleavage to destabilise these transcripts. Interestingly, this mechanism of direct regulation has been shown to modulate the abundance of alternative transcript isoforms. CBP80, Cap Binding Protein 80; DDRNAs, DNA Damage RNAs.

concomitant decrease in the alternative transcript isoform, suggesting the machinery that perform miRNA biogenesis and alternative splicing can compete (Melamed et al., 2013).

In contrast to the previous mechanisms described, a role for Drosha operating as a splicing enhancer was also reported by Havens et al. (2014). The proposed mechanism is that Drosha binds to cassette exons that form pri-miRNA-like structures and promotes their inclusion (Havens et al., 2014). Importantly, this regulation was identified to be independent of Drosha cleavage function.

The Microprocessor's role within alternative splicing and transcription are of particular interest as they highlight two distinct mechanisms. On one hand, it can act as a destabilising unit, negatively regulating RNA species by cleaving dsRNAs hairpins embedded within the transcript. Alternatively, it can act as a scaffold to attract protein co-factors that promote splicing or transcriptional enhancers. The mechanisms that govern these two juxtaposed regulatory functions are currently unknown; however, it is clear that not every pri-miRNA-like structure is processed by the Microprocessor. In order to elucidate this, future work should be geared towards identifying structural RNA elements and protein cofactors that regulate the activity of the Microprocessor complex. Altogether, the findings outlined within this section have highlight relevant non-canonical functions of the Microprocessor complex.

1.1.5 Identification of the direct RNA targets of DGCR8

From the studies outlined above, it is clear the Microprocessor can function outside of miRNA biogenesis. Hitherto, all approaches used to identify RNAs regulated by the Microprocessor were indirect, by comparative transcriptomics of individual cell lines deficient of miRNA biogenesis machinery to identify non-redundant changes in RNA abundancies associated with each deficiency.

In order to identify the RNA targets of the Microprocessor complex Macias et al. (2012) took an alternative approach. On the basis that DGCR8 is the RNA binding component of the Microprocessor, the authors identified the RNAs bound to DGCR8 by high throughput sequencing coupled with immunoprecipitation (HITS-CLIP)(Macias et al., 2012). This technique crosslinks proteins and RNAs that are interacting *in vivo*,

enabling highly stringent purification by immunoprecipitation followed by next generation sequencing. The RNA binding profile of DGCR8 revealed that it does not exclusively bind pri-miRNAs. Instead, DGCR8 binding sites were also found within a plethora of cellular RNAs, such as mRNAs, lncRNAs, small nucleolar RNAs (snoRNAs) and repetitive elements (such as LINE-1 and Alu elements). Subsequent interrogation of mRNAs, lncRNAs and repetitive elements bound by DGCR8 revealed their regulation was dependent on DGCR8 and Drosha (Heras et al., 2013; Macias et al., 2012), which is in agreement with reports of extended non canonical functions of the Microprocessor (Chong et al., 2010; Han et al., 2009; Heras et al., 2013; Kadener et al., 2009; Knuckles et al., 2012). Intriguingly however, the regulation of snoRNAs was DGCR8-dependent but Drosha-independent. These observations indicated that DGCR8 may be forming complexes with other ribonucleases to regulate cellular RNAs, independent of Drosha (Macias et al., 2012).

These novel findings that DGCR8 and Drosha could function independently of each other were corroborated recently in a genetic study in *D.melanogaster* neuronal cell lineage (Luhur et al., 2014). However, it is currently unclear if the Drosha-independent function of DGCR8 uncovered by Macias et al. contributed defects seen in the study from Luhur et al. (2014).

1.2 Part 2: The Ribonucleolytic Eukaryotic Exosome Complex

1.2.1 Introduction

The process of RNA maturation requires a plethora of modifications and processing steps, errors in these lead to the production of aberrant RNAs which are identified by surveillance mechanisms and result in their rapid decay. The exosome complex is the major 3'→5' degradation machinery within the cell and participates in both processing and turnover of RNAs. Originally, components of the human exosome complex were identified as autoantigens that cause the autoimmune disorder polymyositis (PM), scleroderma (scl) or the combined syndrome polymyositis scleroderma (PM/scl) (Gelpi et al., 1990). Later on, the 3'→5' ribonucleolytic activities of the exosome complex were identified and implicated in the processing of 7S ribosomal (r)RNA in yeast (Mitchell and Tollervey, 1996). Thereafter, it was established that the exosome complex is structurally conserved from bacteria to humans, and functions to turnover and/or process a wide range of RNAs both in the nucleus and the cytoplasm. The majority of our understanding of the exosome complexes functions are drawn from studies within yeast, whereas its human counterparts remain to be fully elucidated (for review see Sloan et al., 2012).

1.2.2 The exosome complex functions to degrade and process RNAs

Within yeast, the nuclear exosome complex has been shown to processes 7S rRNA, small nuclear (sn) RNAs, small nucleolar (sno) RNAs and transfer (t) RNAs (Allmang et al., 1999, 2000, van Hoof et al., 2000a, 2000b; Mitchell and Tollervey, 1996; Mitchell et al., 1997; Schneider et al., 2012). Within both yeast and mammals the exosome complex also functions as part of a surveillance network to degrade rRNA, snRNAs, snoRNAs, tRNAs and mRNAs which are erroneously processed, spliced or packaged into RNP complexes (summarised in figure 7) (Lubas et al., 2011; Schilders et al., 2007; Sloan et al., 2013). As part of this pathway, the exosome also degrades by-products from endonucleolytic cleavages, during the release of mature rRNAs. In addition, the eukaryotic exosome complex is also recruited to degrade pervasive RNAPII transcripts (Lubas et al., 2011, 2015; Schneider et al., 2012). Specifically, within yeast, these

transcripts are cryptic upstream transcripts (CUTs) and stable unannotated transcripts (SUTs) (van Dijk et al., 2011; Xu et al., 2009) and in humans these are termed promoter upstream transcripts (PROMPTs)(Preker et al., 2008).

Within the cytoplasm, the exosome operates predominantly as a surveillance mechanism, turning over faulty and regular mRNAs. The majority of mRNA is turned over after translation, and this process involves poly (A) decay followed by exo- and endonucleolysis by ribonucleases which include the exosome complex. In addition, the exosome is also required for the turnover of a number of faulty mRNAs that have no termination codons (NSD; non-stop decay), stalled ribosomes (NGD; no-go decay) or premature termination codons (NMD; non-sense mediated decay)(Houseley et al., 2006).

The cytoplasmic exosome complex also rapidly degrades adenine-uridine-rich mRNAs (AU). These mRNAs contain AU rich elements, commonly within their 3' UTR, which recruit factors that either promote the stabilisation or destabilisation of these mRNAs. In a similar fashion, viral mRNAs are also degraded by the cytoplasmic exosome complex. In humans, the exosome functions in antiviral defence by degrading viral mRNAs which contain zinc-finger antiviral protein (ZAP)-response elements (ZREs). Furthermore, the exosome functions as a scavenger complex and degrades unprotected

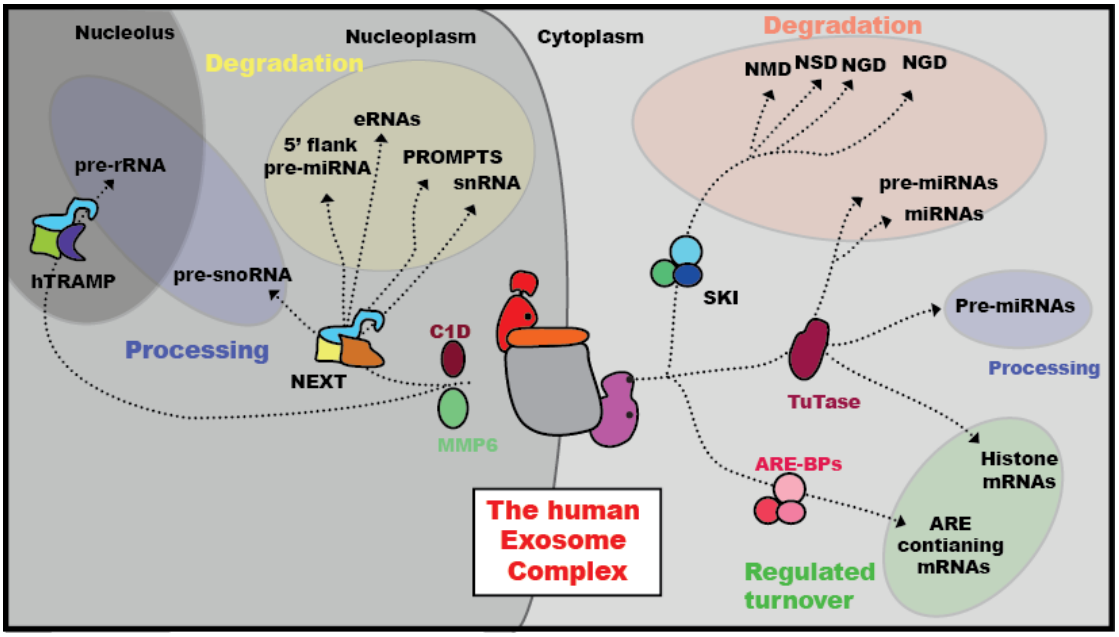


Figure 7 | The functions of the human exosome complex described thus far. Adapted from Lykke-Andersen et al. (2009)

free 3' ends. As such, the 5' products of mRNAs that are destabilised by endonucleolytic cleavage are rapidly recognized and eliminated by the exosome.

Recently, the functions of the cytoplasmic human exosome were also expanded to surveillance of pre-miRNAs, where it polishes 3' ends or degrades aberrant pre-miRNAs (Liu et al., 2014). The polishing of the 3'end of pre-miRNAs with more than 2nt 3' overhangs creates viable Dicer substrates which can therefore be processed into functional mature miRNAs. For these two functions, the exosome requires additional co-factors to stimulate processing or degradation. In this instance, these were identified as the terminal uridylyltransferase (TUTase) which add uracil ribonucleotides at the 3'end of their substrates to stimulate exosome activity (Liu et al., 2014). This activity has also been shown to be important for priming the degradation of histone mRNAs, which is discussed later in this introduction (Mullen and Marzluff, 2008).

1.2.3 The exosome core, a ubiquitous scaffold

All three kingdoms of life have a structurally and functionally similar complexes that perform 3'→5' RNA degradation. The exosome complexes of archaea and eukaryotes are likely to have originated from eubacterial proteins required for RNA decay. Their overall structure is similar and proteins involved display homologous domains which suggests the evolution of the exosome complex pre-dates divergence of the eukaryotes and archaebacterial. What is conserved amongst exosome complex from all kingdoms of life is a barrel-like structure which is wide enough to channel RNA through and is formed by six proteins that are related to the eubacterial RNase PH (PHosphorylsis domain; a 3'→5' exoribonuclease with nucleotidyltransferase domain)(Fig. 8 A).

Within bacteria, the PNPases (polynucleotide phosphorylases) are responsible for RNA decay and form a pseudo-hexameric ring through three PNPase molecules that form a homotrimeric ring of proteins which are homologous to the bacterial RNasePH (Fig. 8 B). However, unlike its RNasePH counterpart, only one member of the dimer retains ribonuclease activity (Symmons et al., 2000). Similarly, the archaeal exosome is composed of six proteins comprising three heterodimers (aRrp41 and aRrp42), which are homologous to the eubacterial RNasePH proteins, however only aRrp41 has retained its

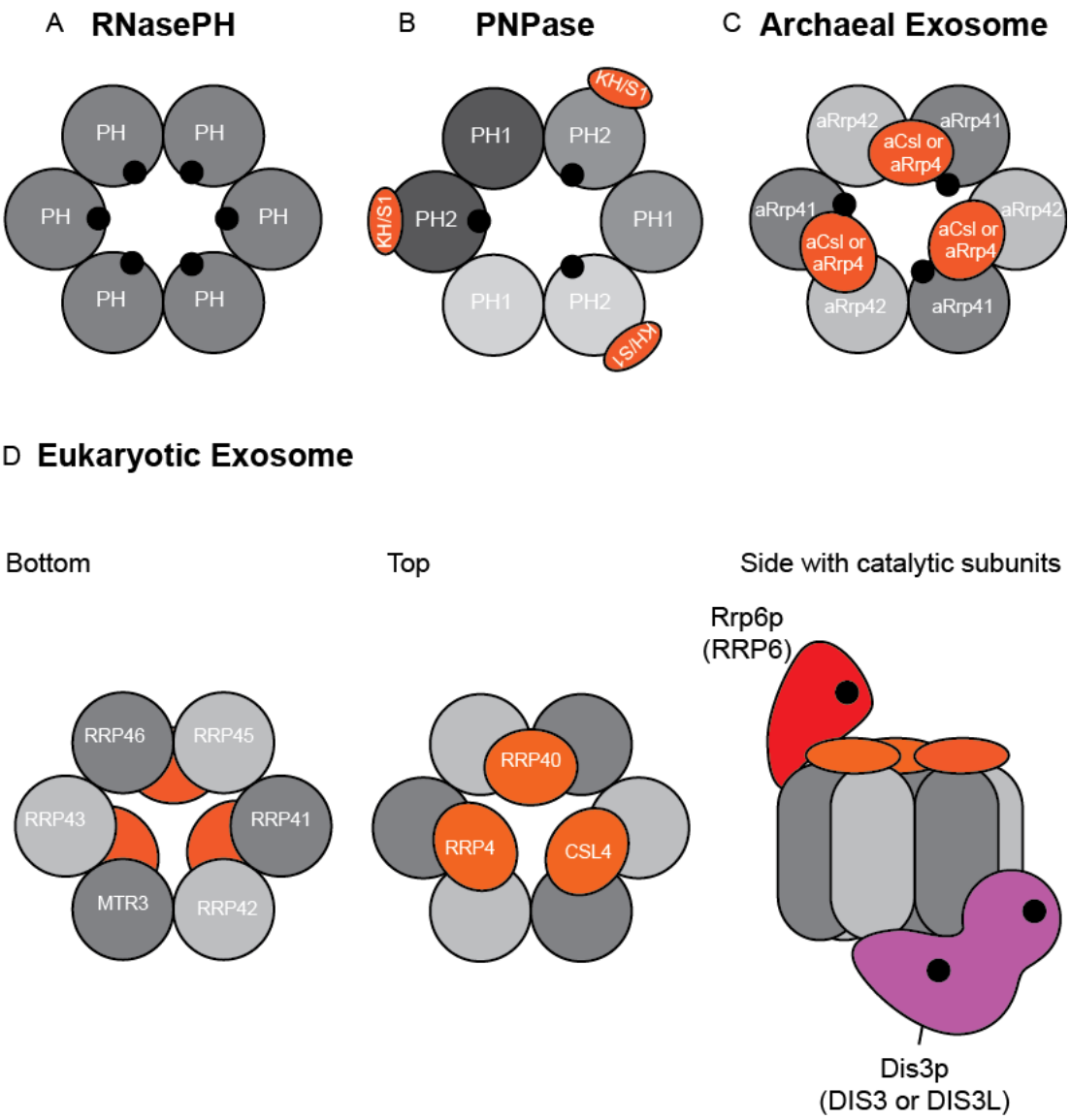


Figure 8 | Exosome core is structurally conserved across a number of different clades. Schematic representation of all structurally conserved RNA degradation complexes from bacteria to eukaryotes. (A) The bacterial RNasePH complex is a homo-hexamer, which forms a barrel structure. All of the components are catalytically active. (B) Another related complex in Bacteria is the PNPase complex, which forms a homotrimeric ring of proteins homologous to RNase PH. Three PH1/PH2 dimers constitute the pseudo-hexameric ring, but only one member of each dimer (PH2) has catalytic activity. In addition, the PH2 proteins contain KH/S1 domains which possess RNA binding activity, and may facilitate PNPase substrate recognition. (C) The Archaeal exosome complex consist of four proteins which are homologous to their Eukaryotic counterparts. The core is comprised of aRrp42/aRrp41 dimers that, like their bacterial PNPase equivalent, for a homotrimeric ring and the aRrp41 subunits have catalytic active. In addition, different combinations of aCsl or aRrp4 proteins that contain KH/S1 domains, cluster around one entrance to the exosome core. The function of these proteins is analogous to the KH/S1 domains possessed by PH2 proteins in the PNPase complex. (D) Displayed is structure of the Eukaryotic exosome complex within yeast; in parenthesis are the human ribonuclease homologues. The Eukaryotic exosome complex consists of 9 distinct proteins which share homology to the subunits of the Archaeal exosome. Six proteins form a heterohexameric barrel structure, while 3 proteins, which again contain KH/S1 domains. Cluster at the opening to the exosome core. The core subunits of the eukaryotic exosome complex are catalytically inert. Interestingly, the eukaryotic exosome associates with additional subunits with ribonuclease activity positioned at opposite entrances to the barrel structure; within yeast this is Dis3p and Rrp6p, whereas in humans RRP6, DIS3 and DIS3L1. DIS3L1 share the same docking region for the human exosome complex. Proteins and domains that contain KH/S1 domains are orange; subunits of the core are shown in various shades of grey; active sites are black dots.

activity (Fig. 8 C). The active sites of aRrp41 are positioned facing into the barrel of the core in order to degrade RNA that is threaded through the channel. In addition, the archaeal exosome is comprised of three extra proteins which form either a homo- or heterotrimer and consist of aRrp4 and/or aCsl4. These proteins are clustered around the opening of the barrel and contain KH and S1 domains (an RNA binding domain originally identified in hnRNPK [KH] and ribosomal protein 1 [S1]) which are important for RNA binding; interestingly, these are also present in the eubacterial PNPases (Fig. 8 B; Table 2).

Table 2 | Components and cofactors of the exosome complex from all domains of life. The a suffix stands for the archaeal version of this protein. Parenthesis contain alternative names. – indicates protein or complex not present in species, or for archeal, the yeast and human homologue. Localisation (Loc) of proteins, can be annotated as: np, nucleoplasmic compartment; no, nucleolar compartment; n, nuclear compartment (no+np); c, cytoplasmic; all, all compartments (no + np + c). Table adapted from Schmid and Jensen (2008).

	Domains	Archaea	Yeast	Loc	Human	Loc
Exosome core	S1 and KH domains	aCsl4	Csl4p	all	CSL4 (EXOSC1)	all
		aRrp4	Rrp4p	all	RRP4 (EXOSC2)	all
		(aRrp4)	Rrp40p	all	RRP40(EXOSC3)	all
	RNasePH	aRrp41	Rrp41p	all	RRP41(hSki6; EXOSC4)	all
		aRrp41	Rrp46p	all	RRP46 (EXOSC5)	all
		aRrp41	Mtr3p	all	MTR3 (EXOSC6)	all
		aRrp42	Rrp42p	all	RRP42 (EXOSC7)	all
		aRrp42	Rrp43p	all	RRP43 (OIP2; EXOSC8)	all
		aRrp42	Rrp45p	all	RRP45 (PmScl-75EXOSC9)	all
Catalytic Subunits	RNase II (exo)	-	Rrp6p	n	hRRP6 (PM/Scl-100;EXOSC10)	all
	RNase D/PIN (exo/endo)	-	Rrp44p (Dis3p)	all	hRRP44(hDIS3)	np+c
	RNase D/PIN (exo/inactive-endo)	-	-	-	hDIS3L1	c
	RNase D(exo)	-	-	-	hDIS3L2	c
Co-factors	Nucleic acid binding	-	Rrp47p (Lrp1p)	n	C1D (hRrp47)	no+np
	RNA binding	-	Mpp6p	n	MPP6	no+np
TRAMP	Poly(A) polymerase	-	Trf4p	n	PAPD5 (hTRF4-2)	no+np
		-	Trf5p	n	PAPD5 (hTRF4-2)	no+np
	RNA binding	-	Air1p	n	ZCCHC7	no
	RNA binding	-	Air2p	n	ZCCHC7	no
	DExH/D Helicase RNA helicase	-	Mtr4p (Dob1p)	n	hMTR4 (SKIV2L2)	no+np
NEXT	RNA binding	-	-	-	RBM7	np
	RNA binding	-	-	-	ZCCHC8	np
	DExH/D Helicase RNA helicase	-	-	-	hMTR4 (SKIV2L2)	no+np

The eukaryotic exosome complex is comprised of nine proteins (Exo-9, within yeast; EXO-9, within humans; Table 3) which share homology to their archaeal and eubacterial counterparts (Fig. 8 D). The crystal structures of the human and yeast exosome complexes identified that the structure of the eukaryotic exosome complex is conserved and consists of two main structures, a core and a cap (Makino et al., 2015; Wasmuth et al., 2014). Six PH-like proteins constitute the exosome core (Rrp41p, Rrp45p, Rrp46p, Rrp43p, Mtr3p, Rrp42p), which forms the highly conserved barrel structure, whilst three proteins form the cap (Rrp4p, Rrp40p, Csl4p), which is located at one entrance of the exosome channel (Fig. 8 D). The components that form the hexameric barrel of the core are proteins homologous to aRrp41 and aRrp42 which form aRrp41/aRrp42-like heterodimers (for homologous archaeal, yeast and human proteins see Table 3). Within humans, these consist of RRP41/RRP45, MTR3/RRP42 and RRP46/RRP43. Whereas three KH/S1 domain proteins, CSL4, RRP4 and RRP40, which are required for exosome stability and substrate recognition, are grouped at the entrance of the exosome complex, as the cap structure (Fig. 8 D & Table 3).

In contrast to the eubacterial and archaeal complexes, the barrel of the eukaryotic exosome has lost its ribonuclease activity (Dziembowski et al., 2007; Liu et al., 2006), and acts as a ubiquitous scaffold for the association of the additional catalytic subunits with ribonucleolytic activity. Furthermore, the barrel structure functions as a channel to direct RNA substrates to the ribonucleases, Dis3p or its human homologues DIS3 or DIS3L1.

1.2.4 Subcellular isoforms of the Eukaryotic exosome complex

Within yeast, the two catalytic components of the exosome, Rrp6p and Dis3p have different localisations patterns; Dis3p is distributed throughout the cell, whereas Rrp6p is nuclear restricted. These different localisation patterns enable the exosome complex to form different subcellular isoforms: within the cytoplasm the core associates with Dis3p, whereas in the nucleus both Dis3p and Rrp6p associate with Exo-9. The ubiquitous association of Dis3p with Exo-9 has led to it being described as the 10th member of the exosome (Exo-10), whereas the nuclear complex is termed Exo-11, as it

Table 3 | Abbreviations used for exosome complex structures

S.cereivisae		Human	
Abbreviations	**Subunit**	**Abbreviations**	**Subunit**
Exo-9	Rrp41p; Rrp45p; Rrp46p; Rrp43p; Mtr3p; Rrp42p; Rrp4p; Rrp40p; Csl4p	EXO-9	RRP41; RRP45; MTR3; RRP42; RRP46; RRP43; CSL4; RP4; RRP40
n/a	n/a	EXO-9 + RRP6	RRP41; RRP45; MTR3; RRP42; RRP46; RRP43; CSL4; RP4; RRP40; RRP6
Exo-10	Rrp41p; Rrp45p; Rrp46p; Rrp43p; Mtr3p; Rrp42p; Rrp4p; Rrp40p; Csl4p; Dis3p	EXO-9 + DIS3	RRP41; RRP45; MTR3; RRP42; RRP46; RRP43; CSL4; RP4; RRP40; DIS3
Exo-11	Rrp41p; Rrp45p; Rrp46p; Rrp43p; Mtr3p; Rrp42p; Rrp4p; Rrp40p; Csl4p; Dis3p; Rrp6p	EXO-9 + DIS3 + RRP6	RRP41; RRP45; MTR3; RRP42; RRP46; RRP43; CSL4; RP4; RRP40; DIS3; RRP6
Exo-12	Rrp41p; Rrp45p; Rrp46p; Rrp43p; Mtr3p; Rrp42p; Rrp4p; Rrp40p; Csl4p; Dis3p; Rrp6p; Rrp47p	EXO-9 + DIS3 + RRP6 + C1D	RRP41; RRP45; MTR3; RRP42; RRP46; RRP43; CSL4; RP4; RRP40; DIS3; RRP6; C1D

contains all core components and associated catalytic subunits (please reference Table 2 & 3 for exosome core abbreviations, homologoues, and subcellular localisation).

Within humans, the picture is more complex as there are 2 additional DIS3 (homolog to Dis3p) family members, DIS3L1 and DIS3L2. Furthermore, all the catalytic subunits have different subcellular localisation patterns (Fig. 9). For instance, RRP6 is highly enriched in the nucleolus, but also distributed throughout the cell; whereas, DIS3 is excluded from the nucleolus but present in the nucleoplasm and cytoplasm; while DIS3L1 and DIS3L2 are exclusively cytoplasmic (Tomecki et al., 2010). Due to the diverse subcellular localisation patterns of the associated ribonucleases, the human exosome complex can give rise to a number of subcellular isoforms. Within the nucleolus, the core exosome (EXO-9) associates with RRP6, whereas within the nucleoplasm EXO-9 associates with both RRP6 and DIS3. In the cytoplasm, DIS3 or DIS3L1 can associate with the EXO-9 and RRP6. The cytoplasmic complexes are thought to be mutually exclusive as DIS3 and DIS3L1 both interact with the same region of the core. Interestingly, DIS3L2 functions independently of the exosome core (Lubas et al., 2013).

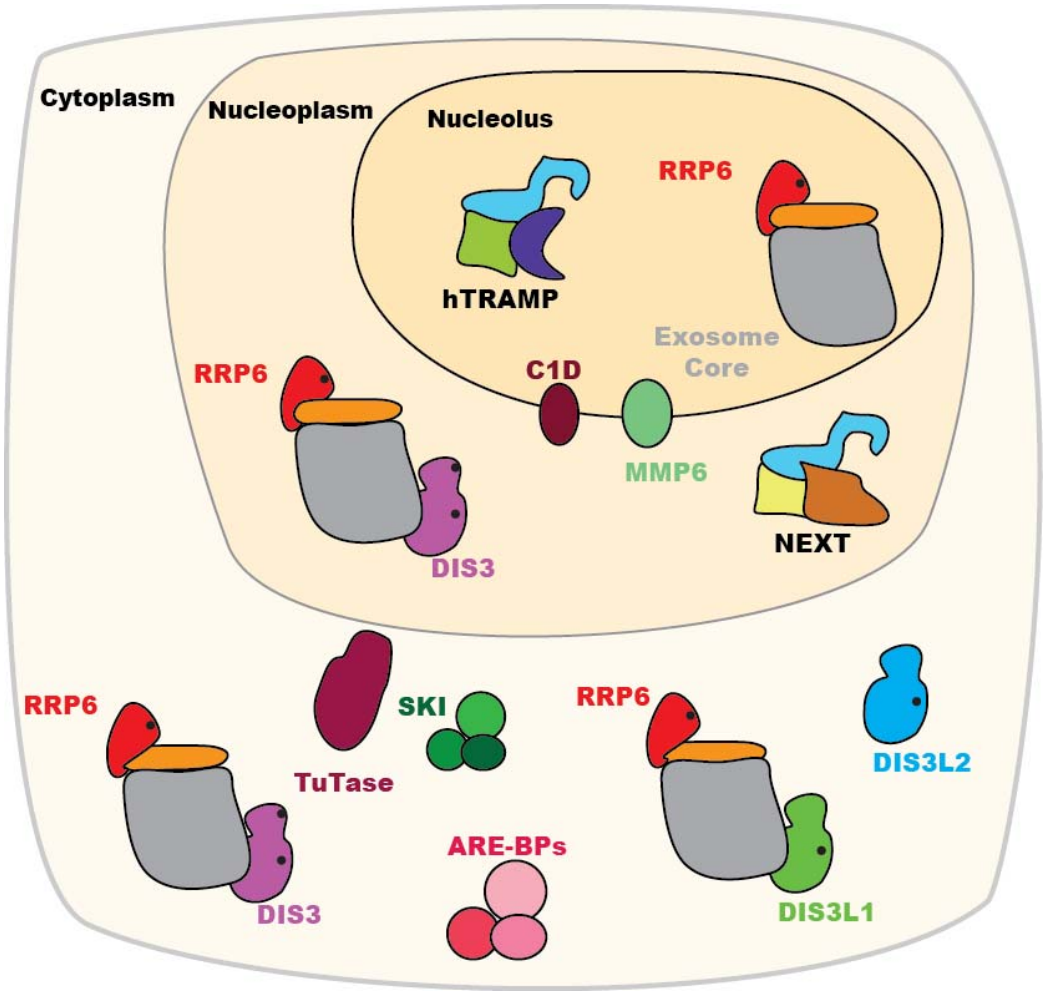


Figure 9 | Subcellular localisation of the human exosome and its associated cofactors and adapters. Schematic representation of the human exosome and its associated ribonucleases, cofactors and adapter complexes. The exosome core (orange and grey) is ubiquitous throughout the cell, whereas its associated ribonucleases are localised throughout different compartments. RRP6 (red) is highly enriched within the nucleolus, but is also present in the nucleoplasm and cytoplasmic compartments; whereas DIS3 (purple) is excluded from the nucleolus and enriched within the nucleoplasm, but also present in the cytoplasm. DIS3L1 (green) and DIS3L2 (blue) are both exclusively cytoplasmic, however, DIS3L2 functions independently of the exosome core. The putative hTRAMP and NEXT complex are restricted to the nucleolar and nucleoplasmic compartments, respectively. The cofactors C1D and MMP6 are nuclear which constitute omnipresent components of the nuclear exosome complex. Within the cytoplasm, the exosome complex is recruited by cofactors or adapter complexes to mRNA to induce their degradation. Those displayed in the schematic are: AU rich element RNA binding proteins (ARE-BPs), which bind AU rich sequences within mRNAs and recruits the exosome; Terminal uridylyl transferases (TUTases), which oligouridylated 3' end of RNAs, such as histone mRNAs, which ultimately results in the recruitment of the exosome complex; or the SKI complex also activates degradation of aberrant mRNAs by recruiting the exosome.

1.2.5 The catalytic subunits of the eukaryotic exosome

1.2.5.1 The DIS3 proteins are members of the RNase II protein family

The Dis3p/DIS3 proteins are catalytic subunits of the exosome can perform both 3'→5' exo- and endonucleolytic processing and conserved from humans to yeast

(Lebreton et al., 2008). Dis3p is positioned at the bottom of the exosome core, in relation to the trimeric cap, and interacts with PH-like subunits Rrp41 and Rrp45 of the exosome core through its N-terminal PIN domain (Fig. 10 A)(Bonneau et al., 2009; Malet et al., 2010; Schneider et al., 2009). Additionally, this domain is also responsible for the endonucleolytic activity of Dis3p, whereas the exonucleolytic activity is contained within the RNB domain (Fig. 10 B). Structural and genetic studies of the yeast exosome have identified that there are three mechanisms of Dis3p-mediated RNA degradation. Firstly, exonucleolysis can occur via a channel-dependent or channel-independent mechanism. Channel-dependent exonucleolytic processing requires delivery of RNA to the active site of Dis3p by threading of substrates through the barrel (for schematic see Fig. 10 D) and given the barrel can only accommodate ssRNA which is 30nt long (Bonneau et al., 2009), this mechanism of processing requires >30nt stretch of ssRNA (Makino et al., 2015). In addition, RNase protection assays of reconstituted Exo-10 identified, two different lengths of substrates were protected by Dis3p. A longer 30-33nt, representing RNAs protected by the Exo-9 and an additional 9-12nt RNA (Bonneau et al., 2009). Protection of these smaller RNA fragments identified that an alternative route to the active site of Dis3p may exist. Subsequent crystallography studies of reconstituted Exo-12 (Exo-10+Rrp6p+Rrp47p) bound to an RNA substrate corroborated these results, revealing Dis3p could bind RNA substrates in a channel-independent manner (Fig. 10 E; see Dis3p exo-) (Makino et al., 2015). Specifically, 12nt were bound within the RNB domain and the neighbouring Rrp45p protein, which is in line with previous crystal structures and RNase protection assays (Bonneau et al., 2009; Lorentzen et al., 2008). To achieve this, there is a change in structural conformation of Dis3p, as the RNB domain was rotated away from opening of the exosome core, whilst still interacting with the core via its PIN domain (Makino et al., 2015). Interestingly, Dis3p has been shown to degrade dsRNA structures without the aid of RNA helicases (Bonneau et al., 2009). The energetic requirements of Dis3p to unwind RNA are produced from the exonucleolytic processing of ssRNA in bursts (Lee et al., 2012).

The mechanism of substrate delivery to the PIN domain of Dis3p for endonucleolyisis is still uncertain (Fig. 10 E; see Dis3p endo-). Temperature sensitive

yeast mutants show that blockage of the exosomes channel reduces both exo- and endoribonucleolytic properties of Dis3p. In addition, exo- and endonucleolytic Dis3p mutants accumulated different intermediates of RNAs usually degraded by the exosome complex (Dziembowski et al., 2007; Schneider et al., 2009, 2012). Furthermore, mutation of the other associated ribonuclease, Rrp6p, were shown to affect Dis3p processivity. In agreement with this, *in vivo* characterisation of the RNA targets of Dis3p identified different profile of substrates associated with the endo- and exoribonucleolytic domains (Schneider et al., 2012). These observations suggest functional processing requires cooperation between the ribonucleases located at alternative ends of the exosome, and that the endonuclease domain may function to rescue blocked channels of the exosome.

Within yeast Dis3p is present in both the nuclear and cytoplasmic compartments and is constitutively associated with the core. In humans however, DIS3 is not ubiquitously present with core. As previously stated the DIS3 family homologues have different subcellular localisation patterns and 3'→5' exonuclease activity. Interestingly, only DIS3 has retained endoribonucleolytic activity. This may be attributed to poor conservation of the DIS3 PIN domain in both DIS3L1 and DIS3L2. Interestingly, the PIN domain of DIS3L2 has diverged so much it can no longer interact with the EXO-9, and functions independently of the core (Fig. 10 B) (Lubas et al., 2013).

1.2.5.2 RRP6 is a member of the RNase D protein family

Rrp6 is a 3'→5' exoribonuclease that associates with the exosome complex. Rrp6 is a member of the RNase D family of proteins which contain a conserved DEDD-Y motif within their active sites and is critical for nucleolytic activity. Within yeast, Rrp6p is the only member of the exosome complex that is non-essential. The C terminal domain, which contains a NLS, interacts with the outside of the exosome core proteins Mtr3p, Rrp43p and Csl4p, and this has been observed in all solved exosome structures to date (Fig. 10 A) (Makino et al., 2015; Wasmuth et al., 2014). The PMC2NT domain is contained at the N terminus and interacts with Rrp47p (C1D in humans), which is

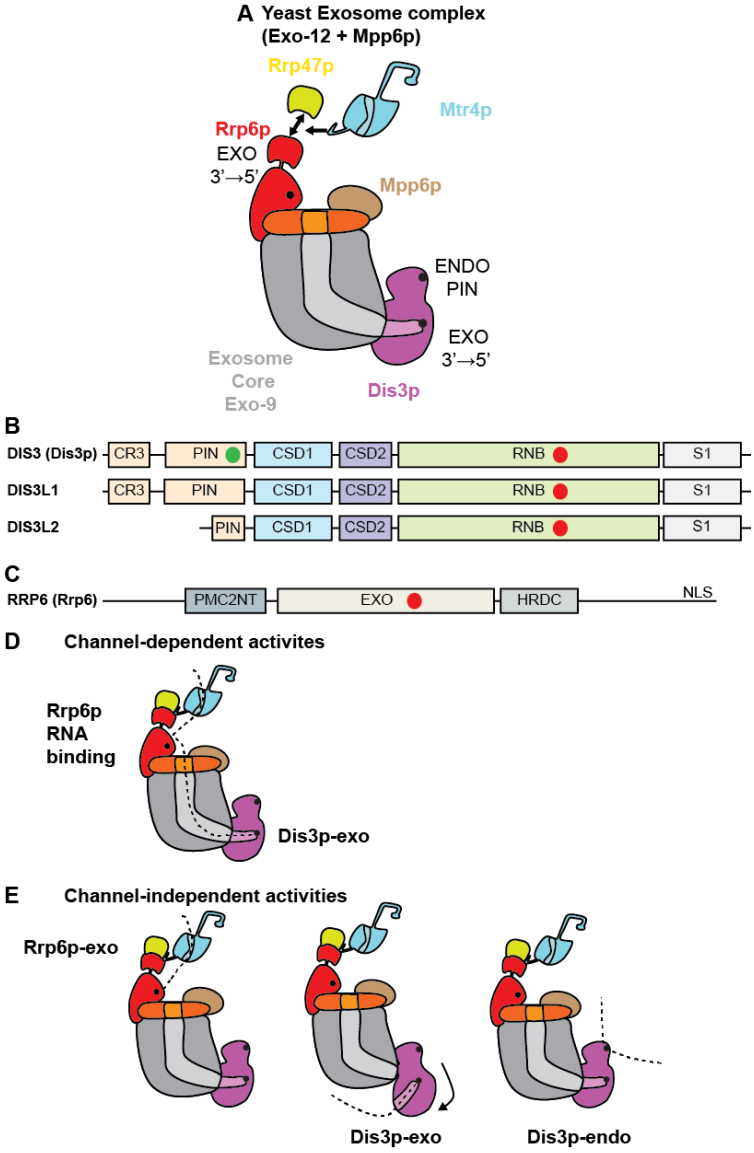


Figure 10 | Schematic DIS3 DIS3L DIS3L2 and RRP6 and the yeast exosome complex and associated proteins started. **(A)**Cartoon of yeast exosome complex (Exo-12 + Mpp6p). Adapted from Makino et al. (2015) & Schuch et al. (2014).Exonuclease and endonuclease domains highlight with black dots. Rrp47 and Rrp6p interact to form a composite binding site for Mtr4p (arrows). Also Mtr4p has a specialised arch domain, and a DexH helicase core which RNA substrates are unwound through. (B,C) Schematic representation of proteins domains within the human exosome associated ribonucleases RRP6, DIS3, DIS3L and DIS3L2. (B) The DIS3 protein family contains; a PIN (PilT N terminus) domain, which harbours the endoribonucleolytic activity of DIS3 (green dot); a CR3 motif also supports the endoribonucleolytic activity of the PIN domain. In addition, DIS3 also contains a RNB domain which contains the RNase II catalytic activity (red dot); and, a S1 domain and a two cold shock domains (CSD1 and CSD2) which support the exoribonucleolytic activity of the RNB domain. Interestingly, DIS3 is the only member of this protein family to have retained endonucleolytic activity, whereas DIS3L1 contains a poorly conserved PIN domains, while DIS3L2 PIN domain is shorter, and as a results DIS3L2 cannot interact with the exosome core and functions independently. (C) RRP6 has an N-terminal domain, PMC2NT, which interacts with C1D; a large EXO domain which houses its catalytic active site (red dot); a HRDC (helicase and RNase D C terminal domain), which has been shown to be important RNA interactions; and, a C terminal NLS. (D,E) Cartoon depicting methods of pathways degradation. (D)In this conformation, Rrp6p threads substrates through the core to Dis3p exoribonuclease active site. (E) Channel-independent binding schematic. Substrates are degraded directly by Rrp6p, without the need for threading through the channel (Rrp6p-exo). Alternatively, Dis3p rotates away from the core, opening its active site up for substrates to be degraded via an channel-independent pathway (Dis3p-exo). Dis3p can also cleave RNA substrates with its PIN domain, which does not require substrates being threaded through the channel.

important for the stability of Rrp6p (Feigenbutz et al., 2013a; Schuch et al., 2014). Furthermore, the HDRC (helicase and RNase D carboxyl terminal domain) and EXO domain contribute to the exonuclease activity of Rrp6p, however, the active site residues DEDD-Y (D313, E315, D371, D440, Y436) are contained within the EXO domain (Januszyk et al., 2011)(Fig. 10 B).

The ribonuclease activity of Rrp6p is described as distributive; as it cleaves, releases and then reengages another ribonucleotide, in a nucleotide by nucleotide nature. The activities of yeast and human RRP6 are subtly different. Although the processivity of both are inhibited by highly structured RNAs, human RRP6 has the propensity to degrade through these dsRNA structures (Januszyk et al., 2011). Furthermore, the active site of human RRP6 is much shallower than its yeast counterpart, which could increase its accessibility to ssRNA (Januszyk et al., 2011).

The crystal structure of the reconstituted yeast and human exosome complexes have been solved by the Conti and Lima group, respectively. The reconstituted structures consist of a yeast exosome complex with ribonucleases and a cofactor (Exo-12; Exo-9 + Dis3p + Rrp6p + Rrp47p) (Makino et al., 2015) and the human structure of EXO-9 + RRP6 (Wasmuth et al., 2014). Both crystallography studies were performed in the presence of an RNA substrate and have provided invaluable insights into the structural organisation of their respective complexes. However, the conclusions drawn from both studies have created controversies, with regards to the mechanism of RNA substrate delivery to the active site of Rrp6. The Conti group suggested the RNA substrate does not pass through Exo-9 to reach the active site of Rrp6p (Fig. 10 E; see Rrp6p) (Makino et al., 2015), whereas the Lima group suggested the RNA substrate must enter the EXO-9 and pass through the S1/KH ring structure (Fig. 10 A; orange structure) and then be diverted out between the cap and PH-like ring structure (Fig. 10 A; grey structure) (Wasmuth et al., 2014). Within the former study, the yeast exosome crystal structure demonstrated an RNA substrate engaged with the active site of Rrp6p, revealing important residues for RNA-Rrp6p contacts. Interestingly, this study also identified Rrp6p can function to bind and direct RNA to the entrance of the core's channel (RNA-binding mode) (Fig. 10 D). Importantly, these functions of Rrp6p are independent of its catalytical activity, and are

in line with previous reports that mutations in Rrp6p affect the processivity of Dis3p (Wasmuth and Lima, 2012; Wasmuth et al., 2014). Finally, the structure of Rrp6p associated with the exosome co-factors Rrp47p and Mtr4p, have identified that Rrp6p and Rrp47p mediate interactions between the Exo-10 and the nuclear adaptor complexes that facilitate decay (Fig. 10 A) (Schuch et al., 2014).

1.2.6 Cofactors of the nuclear exosome complex mediate RNA decay

The associated ribonucleases of the exosome complex may show preferential degradation activities on RNAs with certain ribonucleotide compositions in (Januszyk et al., 2011), however they have no inherent RNA substrate specificity. The associated ribonucleases are scavenger ribonucleases, and they can degrade most unprotected RNAs. Therefore, the *in vivo* activities of the exosome are regulated by associated protein co-factors and RNA adapter proteins which facilitate exosome access and recruitment to the different groups of RNA substrates. Like the exosome associated ribonucleases, the adaptor complexes and co-factors are also localised to different subcellular compartments; which is summarised in Figure 9. Specialised adaptors allows coordinated degradation and/or processing of almost every RNA species present within the cell. Below relevant nuclear adaptors are discussed.

1.2.6.1 C1D and MPP6 are important for RNA processing

Rrp47p (also known as Lrp1p in yeast) is a nucleic acid binding protein and is also important for the processing of nuclear ncRNAs. Originally identified as a protein involved in DNA damage response, Rrp47p also associates with the exosome complex as a nuclear specific cofactor (Mitchell et al., 2003). Both C1D and Rrp47 are small and relatively uncharacterised proteins, with a Sas10/C1D and lysine rich domains respectively located at the N- and C-termini. Most of the exosome associated functions of Rrp47p have been described within yeast. The Sas10/C1D domain of Rrp47p interacts directly with the PMC2NT domain located in the C terminus of Rrp6p (Stead et al., 2007), whereas the C-terminal domains has been shown to interact with components of the box C/D snoRNP (Costello et al., 2011). The interaction between Rrp47p and Rrp6p is mutually beneficial as it is required for the stability of both proteins (Feigenbutz et al.,

2013b). Furthermore, recent structural studies identified that the interaction of Rrp47p-Rrp6p is required to form a composite binding site for Mtr4p (Schuch et al., 2014).

The function of Rrp47p was identified in genetic experiments within yeast cells. Rrp47p deficient cells displayed similar 3'extended forms of snoRNAs, snRNAs, 5.8S rRNA and accumulation 5'-ETS, a rRNA processing by-product (Mitchell et al., 2003), which are also seen in exosome deficient cells (Mitchell and Tollervey, 1996). The defects seen in snoRNA due to depletion of Rrp47p; which has also been shown to interact with the box C/D snoRNP, suggest Rrp47p and the snoRNP may interact to facilitate snoRNA processing (Costello et al., 2011). Within humans, only 5'ETS and 5.8S rRNA are affected by depletion of C1D, identifying a conserved role for Rrp47/C1D in at least rRNA biogenesis (Schilders et al., 2007; Sloan et al., 2013).

Mpp6p (known as MPP6 in humans; M-phase-phosphoprotein) is another cofactor of the yeast nuclear exosome complex that associates directly with the cap structure positioned at the top of the exosome complex (Schuch et al., 2014). Like Rrp47p and Rrp6p, Mpp6p is also required for correct 5.8S rRNA maturation and surveillance and turnover of CUTS (Milligan et al., 2008). Within humans, the functions of MPP6 are conserved, as cell deficient of MPP6 or C1D display similar defects in rRNA processing, identifying an important role for Mpp6p/MPP6 (Schilders et al., 2005).

1.2.6.2 MTR4 is a ubiquitously present cofactor of the nuclear exosome

The exosome cofactor Mtr4p (mRNA transport 4; known as Dob1p in yeast; known as MTR4 or SKIV2L2 in humans) is a 3'→5' RNA helicase that belongs to the DExH/D family. It was originally identified in a genetic screen for yeast mutants that accumulate polyadenylated RNAs within the nucleus (Liang et al., 1996). However, Mtr4p is now regarded as an important co-factor that is omnipresent within the yeast and human nuclear exosome. This intimate association is supported by common processing defects displayed in deficiencies of Mtr4p and/or core exosome subunits (de la Cruz et al., 1998). Specifically, defects in pre-rRNA, pre-snRNA and pre-snoRNA processing were observed in both yeast and mammalian cell lines (van Hoof et al., 2000a; Schilders et al., 2007). All of which suggest the function of Mtr4p is conserved from yeast to humans.

Structural studies of Mtr4p identified a unique arch/KOW domain that is not seen in other DExH/D RNA helicases (Jackson et al., 2010; Weir et al., 2010). The arch/KOW domain is important for the substrate handover from Dis3p to Rrp6p during 7S rRNA processing (Klauer and van Hoof, 2013) and was shown to interact with *in vitro* transcribed tRNAs (Weir et al., 2010), suggesting the arch/KOW domain acts like an arm positioning RNA substrates for subsequent degradation.

Mtr4p can function directly with the exosome complex, however it is also part of all the yeast and human nuclear adapter complexes identified so far (LaCava et al., 2005; Lubas et al., 2011; Vanácová et al., 2005). Mtr4p also binds a composite binding site formed by the interaction between Rrp6p and Rrp47p (Schuch et al., 2014). Therefore Mtr4p may serve as the universal docking protein, mediating interactions between adaptor complexes and the Exo-12.

1.2.6.3 The TRAMP complex

The best characterised co-factor of the nuclear exosome is the *S.cerevisiae* TRAMP (Trf4p/5p-Air1p/2p-Mtr4p-Polyadenylation) complex. The TRAMP complex is associated to the exosome complex and is involved in the nuclear surveillance and turnover of snRNAs, snoRNAs, rRNAs, tRNAs and rRNA (LaCava et al., 2005; Vanácová et al., 2005). The trimeric TRAMP complex is constituted of an RNA helicase, Mtr4p; a non-canonical poly (A) polymerase, either Trf4p or Trf5p; and a zinc knuckle RNA-binding protein, either Air1p or Air2p. The Air1p/2p proteins appear to be functionally redundant, whereas the TRAMP complexes constituted with different Trf4p/5p proteins have different substrate specificities *in vivo* (Egecioglu et al., 2006; Houseley and Tollervey, 2006). Specifically, Trf4p is essential for polyadenylation of all Rnt1p cleavage intermediates, whereas Trf5p only contributes to the polyadenylation of some specific Rnt1p intermediates. Depending on their Trf protein complement, the TRAMP complexes are TRAMP4 and TRAMP5.

Within yeast, the TRAMP complex is recruited to substrates by another exosome adaptor and RNA-binding protein complex called NNS (Nrd1p/Nab3p/Sen1p complex)(Arigo et al., 2006; Thiebaut et al., 2006; Vasiljeva et al., 2008). The NNS complex targets short sequence motifs in nascent RNA transcribed by RNAPII to trigger

transcription termination. This process is also coupled with the transfer of substrates with free 3' ends to the TRAMP complex, which is an important mechanism for the efficient elimination of pervasive RNAPII transcripts within yeast, such as CUTS and SUTs (Tudek et al., 2014).

Recently, the crystal structure of the yeast TRAMP complex was solved; this provided insights into the delivery method of RNA substrates to the exosome complex (Falk et al., 2014). Essentially, the TRAMP complex stimulates exosome activity by unwinding and polyadenylating RNA substrates; and delivers them to Rrp6p and the core exosome. The Air proteins facilitate substrate recognition whilst DExH helicase domain of Mtr4p unwinds structured substrates. Thereafter, the ssRNA exiting the DExH domain is directly channelled towards Rrp6p or Dis3p; or alternatively, it is polyadenylated by Trf4p/5p and then degraded by the exosome. The addition of poly (A) tails by Trf4p/5p differs from those seen in mRNAs as they are shorter and function to destabilise the transcript. The structural organisation of TRAMP bound to the exosome complex remain to be elucidated, however, as Mtr4p binds a composite binding site formed by Rrp6p and Rrp47p (Schuch et al., 2014), it seems likely Mtr4p as part of the TRAMP complex binds here. This would position the TRAMP complex above the exosome channel, and in close proximity to Rrp6p, which would be optimal for its functions.

Sequence homology searches and interaction profiling of the exosome components identified the TRAMP complexes human counterpart (hTRAMP)(Lubas et al., 2011). It is comprised of a zinc knuckle protein, ZCCHC7 (Air1p/2p in yeast); the non-canonical poly (A) polymerase, PAPD5/hTRF4-2 (Trf4p in yeast); and MTR4. The depletion of hTRAMP components caused a decrease in polyadenylated rRNA processing by-products which are usually rapidly turned over by the exosome complex (Lubas et al., 2011). These observations are in line with the nucleolar restriction of the hTRAMP complex, as rRNA processing is a nucleolar process. To date, pre-rRNA intermediates are the only confirmed substrates of the hTRAMP. In addition, the mechanisms of hTRAMP recruitment are unknown, as there is no known homologue of the NNS within humans.

1.2.6.4 The NEXT complex is a mammalian specific adaptor complex

The trimeric NEXT complex is unique to higher eukaryotes and is composed of a zinc finger protein, ZCCHC8; the RNA binding protein, RBM7; and MTR4. Although the NEXT complex shares a common subunit (MTR4), it is distinct from hTRAMP complex, as the NEXT complex is restricted to the nucleoplasm and has no polymerase activity (Lubas et al., 2011). The NEXT complex recruits the exosome complex to a multitude of RNAPII transcripts (Lubas et al., 2015). Furthermore, the NEXT/exosome complex is required for the elimination of 3' extended intronless genes, eRNAs (enhancer RNAs); and pervasive RNAPII transcripts such as, PROMPTs (Promoter UPstream Transcripts)(Hrossova et al., 2015; Lubas et al., 2015). Interestingly, depletion of ZCCHC8 results in 3' extended forms of snoRNAs, suggesting the NEXT complex facilitates exosome recruitment to the 3' intronic sequences flanking pre-snoRNAs.

The NEXT complex was identified by profiling interactions of the exosome subunit RRP6 and cofactor MTR4 (Lubas et al., 2011). This study also identified a number other proteins which were highly enriched in both RRP6 and MTR4 purifications, indicating an intimate association with the exosome. These proteins included a number of uncharacterised zinc knuckle proteins; components of the Cap binding complex (CBC); the CBC-associated arsenic-resistance 2 (ARS2) and ZC3H18/NHN1 proteins. All of these proteins and the NEXT complex are functionally linked and constitute a larger complex termed the CBCA-NEXT complex (Cap binding complex, ARS2, ZC3H18/NHN1, NEXT) (Andersen et al., 2013). A distinctive feature of RNAPII transcripts is their 5' cap which is bound by the CBC. Interestingly, it was found the depletion of components of the CBC decreased association of the NEXT complex to snRNAs and replication dependent histone (RDH) mRNAs, which are short intronless genes produced by RNAPII (Lubas et al., 2015). Following these observations, it was concluded the CBC recruits the NEXT complex to RNAPII transcripts in order to facilitate 3' end maturation or turnover by the exosome complex (Andersen et al., 2013; Hallais et al., 2013; Lubas et al., 2015). A common function of both the TRAMP and NEXT complexes is that they are involved in facilitating exosomal processing and complete

elimination of RNAs. However, the mechanisms that directs exosome-mediated processing versus complete degradation of RNA substrates requires further research.

1.3 Part 3: RNA substrates of the exosome complex relevant to this thesis

1.3.1 Introduction

As stated previously, genetic studies within yeast have uncovered a number of RNA species that are processed or degraded by the exosome complex. For the purpose of this thesis, relevant exosome processed RNAs will be discussed within this section.

1.3.2 snoRNAs

SnoRNAs (small nucleolar RNAs) are a highly abundant and stable class of RNAs and are on average 60nt-150nt in length. These RNAs are a component of a ribonucleoparticle called the snoRNP, which acts as an antisense guide to base pair with target RNAs guiding chemical modifications to rRNAs, tRNAs and snRNAs and facilitating the processing of rRNA. As their name suggests, snoRNAs are highly enriched within nucleoli, and can also be subdivided into two major groups depending on conserved motifs harboured within their mature sequence. The box C/D snoRNAs (also known as SNORD), contains a C box (RUGAUGA) and a D box (CUGA) near the 5' and 3' termini of the snoRNA, respectively (Fig. 11 A)(Huang et al., 1992). The box C/D snoRNAs also contain C' and D' boxes which contain less stringently conserved C and D box motifs. The box H/ACA snoRNAs (also known as SNORA), contain a H box (ANANNA) and a ACA box (ACA) which are usually harboured within single stranded regions of the snoRNA (Fig. 11 B)(Balakin et al., 1996). SNORD and SNORA guide 2'O-ribose methylation and pseudouridylation, respectively by base-pairing with the target position in the rRNA (Ni et al., 1997). Additionally, another group of snoRNAs, scaRNAs (small Cajal RNAs; also known as SCARNAs), which are named after their localisation in the Cajal bodies, comprise both box C/D and box H/ACA snoRNAs (Darzacq et al., 2002).

1.3.2.1 Pre-snoRNA processing

Within eukaryotic cells, there are three different approaches to synthesise snoRNAs (for review see Filipowicz and Pogačić, 2002). Within yeast, the majority of

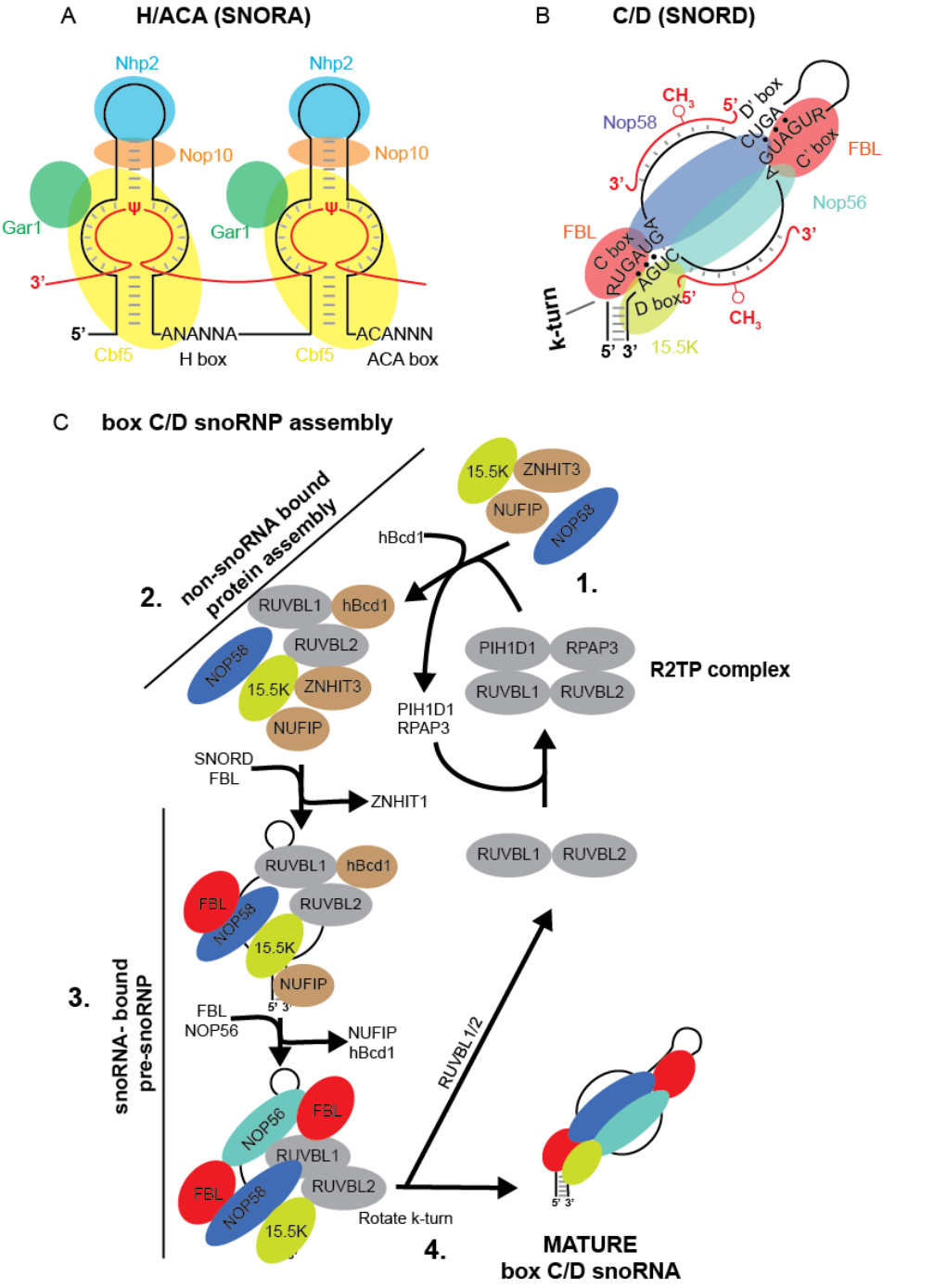


Figure 11 | Biogenesis of the snoRNP. (A,B) Cartoon of mature and functional box H/ACA (A) or box C/D (B) with respective snoRNP bound. (C) Schematic of C/D box snoRNP assembly. 1) Binding of the R2TP complex to 15.5K, NUFIP, ZNHIT3 and NOP58, ejects PIH1D and RPAP3; and also recruits hBcd1. This complex is the proteins assembly for C/D box RNAs and is not associated to the snoRNA yet. 2) The non-snoRNA bound protein assembly is then recruited to the snoRNA. At this point, FBL is recruited and displaces ZNHIT1. 3) Next, another FBL molecule and NOP56 are recruited to the snoRNA, which results in the displacement of NUFIP and hBcD1. 4) During the final step of the box C/D snoRNP assembly the 15.5k subunit of the now snoRNA-bound pre-snoRNA, binds the kink-turn and causes a conformational change in the snoRNA, which also ejects RUVBL1 and RUVBL2 from the complex, resulting in a mature and functional C/D snoRNP. Figure adapted from Bizarro et al. (2014)

Table 4 | snoRNP components. Table listing all Human, Archaeal and Yeast proteins required for a mature snoRNP assembly for both box C/D and box H/ACA RNAs. This table was adapted from Lui and Lowe (2013).

	Humans	**Yeast**	**Archaea**	**Function**
C/D box				
	15.5K	Snu13p	L7Ae	Binds kink-turn formed by C/D motifs
	FBL	Nop1p	Fibrillarin	Methyltransferase
	NOP56	Nop56p	Nop56/Nop58	Recruits FBL
mature snoRNP	NOP58	Nop58p (Nop5p)	Nop56/Nop58	Recruits FBL
	hBcd1	Bcd1		Accumulation of C/D box RNAs, couples transcription and processing of snoRNAs
	TIP48 (RUVBL2)	TIP48p		Bridges interactions between 15.5K and the NOP56/58 prtoeins
	TIP49 (RUVBL1)	TIP49p		Bridges interactions between 15.5K and the NOP56/58 prtoeins
pre-snoRNP	NUFIP			Regulates TIP48 and TIP49 interactions with core C/D box RNA proteins
	PIH1D1			C/D box RNA assembly
	RPAP3			
protein assembly	ZNHIT3			
Non-assembly factor	IBP160			Couples pre-mRNA splicing and snoRNA assembly
H/ACA box				
	NHP2	Nhp2p	L7Ae	Binds to kink-turn
	Dyskerin (NAP57)	Cbf5p	Cbf5	Pseudouridine synthase
	GAR1	Gar1p	Gar1	Target RNA turnover
mature snoRNP	Nop10	Nop10p	Nop10	Interaction with NHP2 and core RNA
pre-snoRNP	NAF1	Naf1p		accumulation of H/ACA box RNAs; swaps with GAR1

snoRNAs are transcribed from their own promoter, but can also be transcribed as a polycistronic unit containing clusters of snoRNAs. In the latter case, snoRNAs are processed from polycistronic transcripts by the RNase III enzyme, Rnt1p (Chanfreau et al., 1998). Thereafter 5' non-snoRNA sequences are removed by the 5'→3' exonuclease Rat1p and Xrn1p (XRN1 in humans)(Petfalski et al., 1998); while the TRAMP complex stimulates exosome-mediated resection of the 3' non-snoRNA sequences (Grzechnik and Kufel, 2008). Mature snoRNAs are protected from complete degradation by association to the core factors of the snoRNP, which defines the 5' and 3' ends of the mature snoRNA. Alternatively, snoRNAs can also be embedded within introns, and their release is facilitated by splicing and lariat debranching, which is the case for most human snoRNAs,

and is in contrast to yeast (Dieci et al., 2009). Interestingly, intronic snoRNAs are housed within introns of genes associated with ribosomal production. Therefore it seems the majority of mammalian snoRNA expression is coupled to expression of rRNA biogenesis proteins, which results in accumulation of factors required for rRNA biogenesis and processing in equivalent amounts. Although this process is quite well defined in yeast, the human biosynthesis pathway requires description and more detailed study. So far, within mammalian cells, the NEXT complex has been identified to target 3' extended snoRNAs to the exosome complex (Lubas et al., 2015). In this same study, snoRNAs with 3' extensions were shown to be stabilised by co-depletion of the exonucleases associated with the human exosome complex, RRP6/DIS3/DIS3L1; however, the individual contributions of these exonucleases to snoRNA processing remain to be elucidated.

1.3.2.2 The snoRNP and its assembly

The assembly of both box H/ACA and box C/D snoRNPs begins as a protein only complex (the pre-snoRNP is functionally inactive)(for review see Matera et al., 2007). For a summary of the proteins involved please refer to Table 4. The H/ACA pre-snoRNP consists of NOP10, NHP2, DKC (Dyskerin; known as Centromere binding factor 5 [Cbf5p]); and the assembly and exchange factor NAF1 (nuclear assembly factor 1) (Fig. 11). The H/ACA box snoRNP guides DKC (a pseudouridine synthase) to chemically modify specific ribonucleotides. Chromatin immunoprecipitation (ChIP) experiments identified these components at actively transcribed H/ACA genes (Ballarino et al., 2005; Darzacq et al., 2006; Richard et al., 2005), which means that these components associated to the pre-snoRNA in a co-transcriptional manner. During the transcription of box H/ACA snoRNAs, the pre-snoRNP is associated with both the C terminal domain (CTD) of RNAPII, mediated by NAF1, and the H/ACA box nascent RNA (Fig. 11 D). During maturation, the snoRNP undergoes dynamic remodelling, as GAR1 binds DKC and replaces NAF1, therefore the active and mature snoRNP contains GAR1, DKC, NOP10 and NHP2. DKC is a pseudouridine synthase and binds the ACA box, anchoring itself and its protein interactors to the H/ACA RNA, and also positioning the antisense boxes near the catalytic active sites of DKC. Additionally, NHP2 binds the apical stem and interacts with NOP10.

Similarly, the assembly of the C/D box snoRNP is also co-transcriptional and is intimately linked to pre-mRNA splicing. These two processes are linked by the splicing factor IBP60, which binds upstream of the box C/D snoRNA and activates C/D RNP assembly. Again, the pre-snoRNP is initially a protein only complex, the components of which were recently identified by interaction profiling experiments (Fig. 11 C) (Bizarro et al., 2014). The pre-snoRNP associates with the box C/D RNA as a complex of proteins including hBcd1 (human box C/D 1), NUFIP, ZNHIT3 and the R2TP complex. Importantly, these proteins are excluded from the mature box C/D snoRNA, whereas FBL (Fibrillarin; the 2'O methyltransferase) and NOP56 are recruited. The mature snoRNP consists of 15.5K (L7Ae homologue), NOP58, NOP56 and FBL. The 15.5K protein binds the k-turn (kink-turn) at the basal stem whereas NOP56 and NOP58 bind the C/D and C'/D' boxes, respectively. Moreover, NOP56 and NOP58 heterodimerise forming a protein link between the box domains, whereas FBL binds and guides the modification of those RNAs base paired with the guide elements.

1.3.2.3 Other snoRNA family members of interest

The box H/ACA motifs within SNORAs are crucial for the recruitment and the subcellular localisation of the functional snoRNP. Interestingly, the box H/ACA motif is also present in the RNA component (TR; telomerase RNA) of the telomerase reverse transcriptase enzyme (TERT), which is essential for maintenance of telomere length (Pogacić et al., 2000). Human TR contains an antisense template of AUCCCAAUC which primes telomere extension and an H/ACA motif within the 3' end which contains a consensus sequence important for Cajal body retention. Intriguingly, the box H/ACA domain is only present in vertebrate TR and as such, its likely TR 3' end processing is performed by the same 3'→5' exonucleases that process box H/ACA snoRNAs, as these two RNAs bind a similar complement of snoRNPs (Dez et al., 2001).

1.3.3 tRNAs

As one of the most abundant small ncRNAs within the cell, tRNAs are essential for the conversion of genomically encoded genes to polypeptide products. Transfer RNAs (tRNAs) are recruited to the ribosome during protein synthesis to guide

specific amino acids to the elongating polypeptide chain by base-pairing with codons present in the mRNA being translated.

1.3.3.1 tRNA biogenesis

tRNAs are generally 60-80nt long and are transcribed by RNAPIII from their own promoter. Transcription is terminated when RNAPIII meets a low complexity T-rich sequence (4> T in a row) (Cozzarelli et al., 1983). Typically transcription start and stop site do not coincided with the 5' and 3' ends of the mature RNA, therefore these pre-tRNAs, like many other ncRNAs, require post-transcriptional processing. The major pathway to remove 5' and 3' trailer sequences requires sequential cleavages by the endoribonucleases RNase P to remove the 5' leader sequence; and RNaseZ (tRNAseZ) to remove the 3' trailer sequence. Thereafter the CCA-adding enzyme ligates a non-genomic CCA to the 3' end of the tRNA, which is important for tRNA recognition during translation. However, recent examination of 3' formation of tRNAs in the absence of RNaseZ identified tRNA with heterogeneous 3' ends, suggesting other ribonuclease maybe involved in the tRNA processing. Further analysis identified the RNase D family of exoribonucleases, contribute to the formation of the 3' ends of tRNAs (Skowronek et al., 2013). Specifically, the redundant family of Rex1/2/3 protein were identified to process tRNAs in addition to Rrp6p. The multiple processing steps of tRNAs are dynamically localised within the cell. RNAPIII transcription and 5' trailer removal occurs within the nucleolus, whereas 3' trailer removal is nucleoplasmic (for review see Maraia and Lamichhane, 2011).

1.3.3.2 Hypomodified tRNAs are degraded by the exosome in yeast

tRNAs are highly modified during their biogenesis which is crucial for their correct folding. The biogenesis of tRNAs is surveyed by factors which sense aberrant tRNAs, such as the TRAMP complex, and target them for degradation. This complex targets hypomodified tRNAs to the exosome complex for their degradation, because they are probably misfolded due to the lack of modifications (Schneider et al., 2007; Vanácová et al., 2005). However, it is currently unknown if these functions of the yeast TRAMP complex are conserved in humans or which surveillance mechanisms are used for faulty tRNAs.

1.3.4 rRNA

Ribosomes facilitate the transfer of genetic encoded information, into polypeptide chains and rRNAs constitute an integral component of this major RNP complex. The major substrate of the exosome complex is rRNA. The mechanisms of rRNA biogenesis have been best characterised within yeast, where 3 mature species of rRNA (18S, 5.8S and 25S) are transcribed by RNAPI as one long 35S unit. In humans, the 3 mature rRNA species are 18S, 5.8S and 28S, which are also transcribed by RNAPI as one long 45S unit. The rDNA operon is organised as follows: 5' ETS, 18S, ITS-1, 5.8S, ITS-2, 25/28S and 3'ETS (ETS, external transcribed spacer; ITS, internal transcribed spacer) (Fig. 12 A). ITS-1 and -2 separate the mature rRNAs, whereas ETS flanks the 5' and 3' regions of 18S and 25/28S, respectively. Release of mature rRNA sequences is catalysed by multiple endo- and exonucleolytic cleavages with coordinated turnover of some of these intervening sequences (summarised in Fig. 12 A).

1.3.4.1 The yeast exosome processes 7S rRNA precursors

The best characterised function of the exosome complex within rRNA biogenesis is the processing of 7S to 6S rRNA during 5.8S rRNA processing (Allmang et al., 1999, 2000). Temperature sensitive (ts) mutants of components of the exosome complex grown at non-permissive temperature display multiple rRNA processing defects. Specifically, 7S rRNA accumulates in exosome deficient cells; indicating that it's processing is blocked. Interestingly, comparative analyses of the cells deficient in Rrp6p or other exosomal components indicated Rrp6p and Dis3p have distinct functions during 7S rRNA processing. Essentially, deficiency of Exo-9 components and Dis3p caused accumulation of 7S (a longer precursor of 6S), while Rrp6p deficiency resulted in accumulation of 5.8S + 30nt (a shorter precursor of 6S) (Fig. 12 B). Interestingly, metabolic labelling experiments from induced depletion of core components (Rrp41p or Rrp45p) from strains already lacking Rrp6p resulted in a progressive decrease in 5.8S+30nt species (Allmang et al., 1999). These epistatic interactions between Rrp41p/Rrp45p and Rrp6p indicated that Exo-10 is important for the processing of 7S to 5.8S+30nt, thereafter Rrp6p takes over processing until the final ~8nt.

1.3.4.2 The functions of the human exosome are conserved within rRNA processing

With the majority of rRNA biogenesis studies performed within yeast, our current understanding of human rRNA biogenesis is less detailed (Fig. 12 C). However, it has been ascertained that the role of RRP6 in 7S rRNA processing is conserved from yeast to humans (Schilders et al., 2007; Sloan et al., 2013). In addition, it has been demonstrated that RRP6 and MTR4 are required for the processing of 18S rRNA intermediates (Sloan et al., 2013). Furthermore, overexpression of a catalytic mutant of RRP6 stabilised specific sections of the 5'ETS which are rapidly turned over in wild type cells. Taken together,

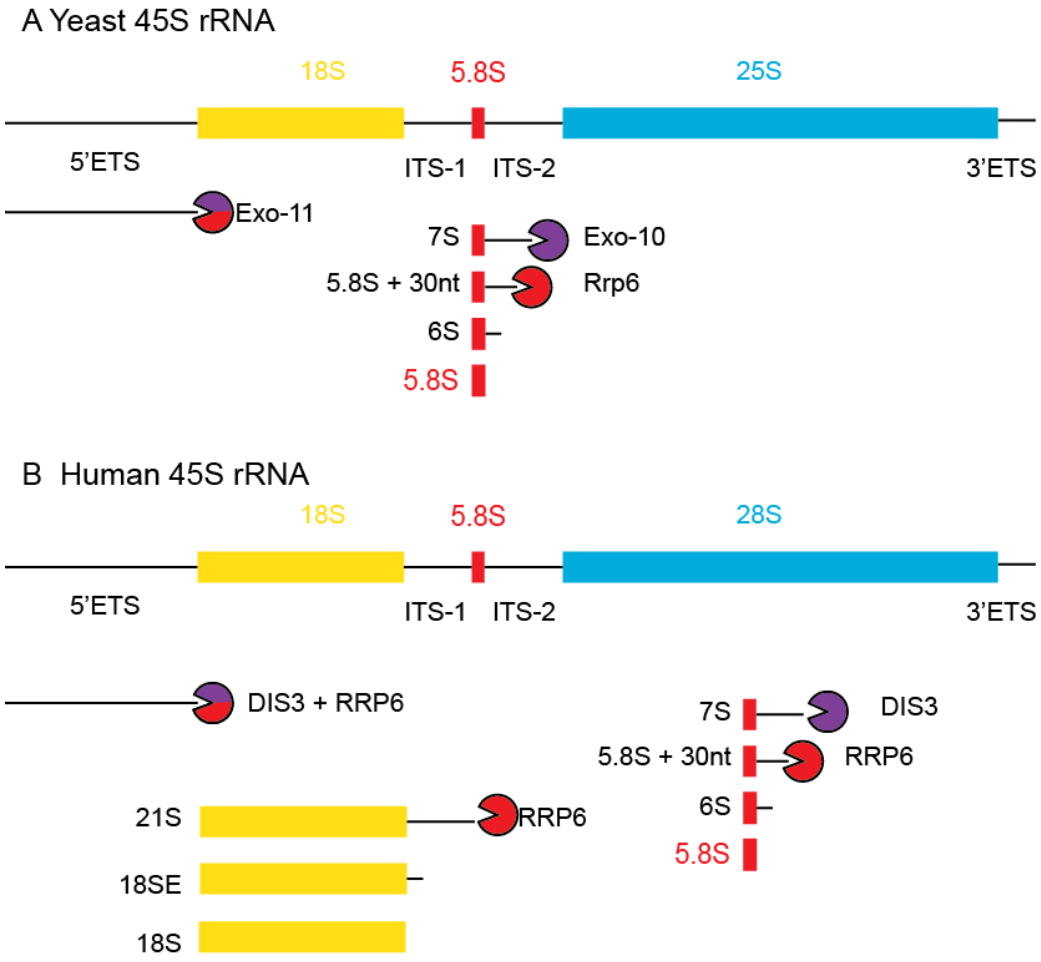


Figure 12 | The exosomes role within rRNA biogenesis. (A,B) A cartoon of the exosomes functions within yeast and human rRNA biogenesis. (A) Exo-10 is required for the processing of 7S to 5.8S + 30nt; thereafter Rrp6 processes 5.8S + 30nt to 6S. Exo-11 is required for the turnover of 5'ETS. (B) In humans these processes are conserved. Additionally RRP6 is required for the processing of 21S to 18 SE

these results have identified three important functions of RRP6 in rRNA biogenesis within humans.

1.3.5 Histone mRNAs

Histone proteins package chromosomal DNA into chromatin, and have a crucial role in gene regulation. Histone proteins can be classified into two groups: replication-dependent histones (RDH) and replication-independent histones (RIH). Like canonical mRNAs, RIH are stabilised by polyadenylation. In contrast, RDH mRNAs are polyadenylation independent and are stabilised by binding SLBP (stem loop binding protein) to a conserved hairpin within their 3'UTR (Battle and Doudna, 2001). During S-phase, there is an increased requirement for histones to package newly synthesised DNA. Accordingly, RDH mRNA transcription and protein expression is upregulated during G1-S phases, as well as SLBP levels (for review see Marzluff et al., 2008).

1.3.5.1 Histone mRNA turnover is activated by TUTases and mediated by the exosome

After DNA synthesis is complete at the end of S-phase, RDH mRNAs are rapidly turned over to inhibit further translation of histone proteins. This is achieved by phosphorylation of SLBP protein by CycA and CDK1, which leads to SLBP degradation (Koseoglu et al., 2008). Subsequently, the unbound hairpin within the 3'UTR of the histone mRNA is oligouridylated by the TUTases (terminal uridylyl transferases), which creates a viable landing pad for Lsm1-7 to bind, which in turn recruits 5' decapping enzymes and the exosome complex to initiate their decay (Hoefig et al., 2013; Mullen and Marzluff, 2008; Slevin et al., 2014).

1.4 Part 4: Identification of direct RNA targets by iCLIP

1.4.1 Introduction

The stability, biogenesis and degradation of all RNAs require dynamic protein:RNA interactions. In order to gain positional data of RNA binding proteins (RBP) interactions with RNA *in vivo*, the crosslinking immunoprecipitation coupled with high throughput sequencing (HITS-CLIP or CLIP-seq) technique was established (Ule et al., 2005). Essentially, CLIP allows the identification of *in vivo* protein:RNA interactions by combining UV cross-linking and immunoprecipitation. UV irradiation covalently cross-links RNA-binding proteins to their RNA substrates which allows a subsequent highly stringent purification. Next, isolated RNAs are processed into libraries for next generation sequencing. The mapping of these sequences generates a map of RNA binding sites and also identify binding motifs for the RBP under study. This approach differs from previous techniques as it offers an unbiased approach towards identification of the direct RNA targets of RBPs *in vivo*.

Since its initial publication, the CLIP technique has been modified to improve cross-linking efficiency, RNA recovery and has been adapted to work within tissues and model organisms(Broughton and Pasquinelli, 2013; Hafner et al., 2010; Konig et al., 2011; Sugimoto et al., 2015). One derivative technique, is the so-coined individual nucleotide resolution CLIP derivative (iCLIP) (Konig et al., 2011). As the name suggests, iCLIP can recover protein:RNA interaction at a nucleotide resolution (Fig. 13). In addition it is also reported to recover a higher percentage of cross-linked RNAs (Sugimoto et al., 2012). The iCLIP technique has been used to successfully identify the direct RNA targets of a series of RBPs, and within this study, RRP6.

1.4.2 Brief overview of protocol

First, living cells are irradiated with 254nm UV light which cross-links interacting protein and RNAs. Next, cells are collected and extract are partially degraded with either a high or low concentration of RNase. Following this, protein:RNA complexes are purified using specific antibodies to the protein understudy which are immobilised on

beads. In addition, a negative control immunoprecipitation (IP) is performed using either IgGs or using the epitope specific antibodies, for endogenous or epitope tagged protein purifications, respectively. The RNA from the protein:RNA complexes immobilised on beads is dephosphorylated and the 3' linkers are ligated. These RNAs are next radiolabelled for visualisation purposes and separated using SDS-PAGE. This step removes free RNA which is not cross-linked to proteins from your complex of interest. Furthermore, the samples that were initially partially degraded with high and low

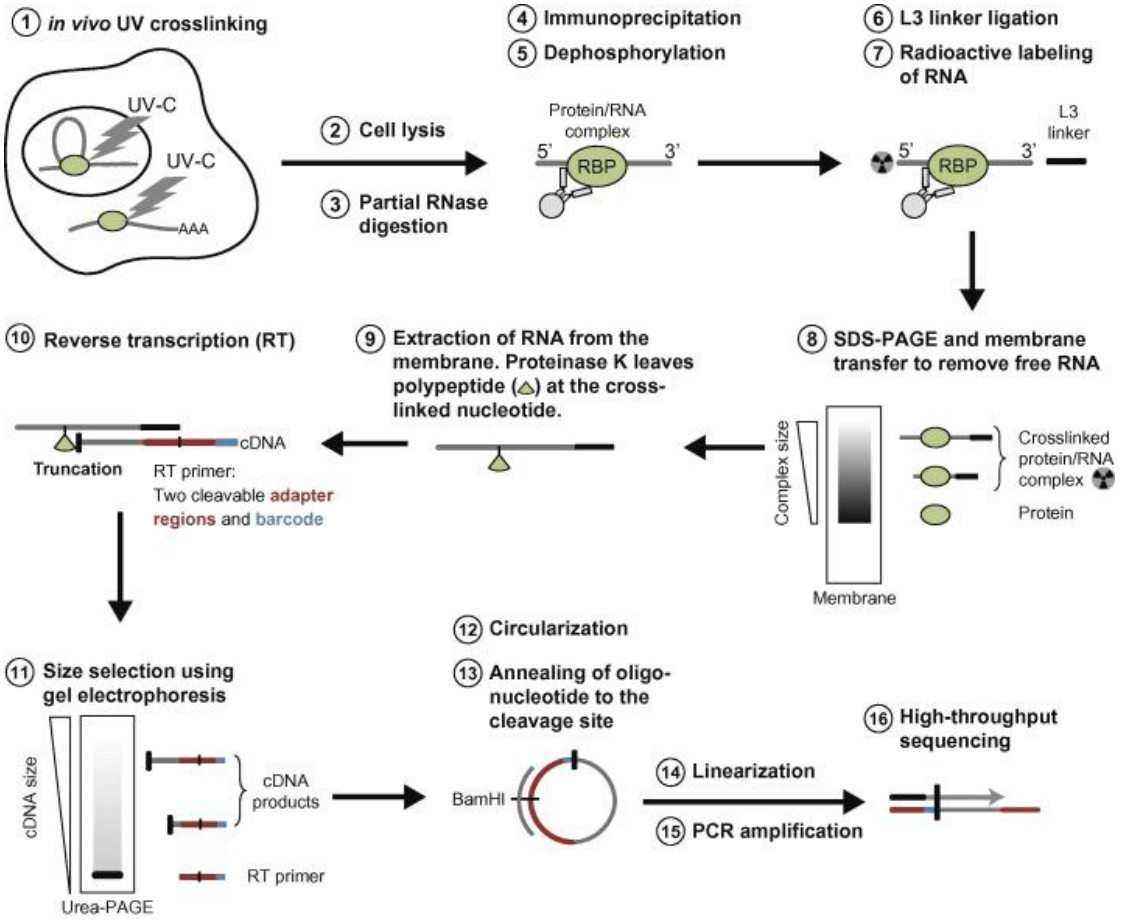


Figure 13 | iCLIP protocol. Schematic of iCLIP protocol to identify protein-RNA interactions at nucleotide resolution. Figure was originally published in Konig et al. (2011). (1) Cells are UV-irradiated to induce protein-RNA crosslinks. After cell lysis (2), RNAs are partially degraded with RNases (3). Next, the RBP:RNA complexes are IPed (4) and RNAs are dephosphorylated (5), and RNA adapters are ligated to 3' end (6), whilst 5' ends are radioactively labelled (7) to visualise RNA:Protein complex after SDS-PAGE (8), which removes free RNAs. Protein:RNA complexes are isolated from above the normal migration of uncrosslinked proteins, which can be detected as a smear of signal (9). Isolated Protein:RNA complexes are treated with proteases, however, small residual polypeptide chains remain crosslinked to the RNA. RNA is then RT (10) which also adds a barcode and 5' adapter sequence in reverse orientation. cDNAs are then separated and size selected via SDS-PAGE (11) to remove RT primers . cDNAs are then circularised by ligating the 5' end of the cDNA to the 3' end of the 5' adapter sequence added during RT (12). Thereafter, the cDNAs are incubated with oligos complementary to a restriction enzyme site in the adapter, which creates a dsDNA substrate (13) which is cleaved by a restriction enzyme (14). (15) cDNAs are amplified by PCR which also adds Illumina adapters. Thereafter, cDNAs are mixed for multiplex sequencing (16).

concentrations of RNase can now be visualised by differential migration of those complexes containing either a short homogenous population of RNAs with faster migration (high RNase) or a longer heterogeneous population of RNAs that retards migration (low RNase). Protein:RNA complexes are then isolated from low RNase samples, whereas high RNase samples are used as a control for immunoprecipitation specificity and purification stringency (explained below). Isolated RNAs are then treated with proteases to digest cross-linked peptides, to synthesize cDNA by using reverse transcription (RT) primers encoding iCLIP barcodes and adapters. The iCLIP adapter is unique from other CLIP protocols as it contains both a 3' and 5' adapter sequence separated by a restriction enzyme site. Next, cDNAs are size selected using gel electrophoresis in order to remove shorter RT primer only sequences. Following this, the isolated cDNAs are circularised, ligating the 3' adapter sequence to the 5' end of the cDNA. Finally, to allow amplification and addition of Illumina adapters for high-throughput sequencing, the circularised cDNAs are linearised by adding an oligo complementary to the restriction site harboured between the 5' and 3' adapter, creating a dsDNA region susceptible for restriction enzyme digestion.

1.4.3 Key differences between CLIP and iCLIP techniques

The iCLIP protocol differs from the original CLIP protocol in that only one intermolecular RNA ligation step is required to introduce the 3' and 5' adapters in order to sequence the bound RNAs (Fig.14). The RNA ligation steps are the least efficient step of these protocols, representing the critical point where bona fide recovered RNAs may be lost. Replacing this with a more efficient intramolecular cDNA circularisation step improves their recovery. In addition, primer extension studies have suggested that residual cross-linked amino acids to the RNA substrates, that are irremovable by proteases, such as Proteinase K, can block reverse transcription during library preparation (Sugimoto et al., 2012; Urlaub et al., 2002). These studies found that only 20% of cDNAs read-through the cross-link site meaning that up to 80% of cDNA transcripts can be truncated at this site (Sugimoto et al., 2012). Therefore, these truncated cDNAs would not be recovered in the original CLIP protocol because both adapter sequences are required to be incorporated

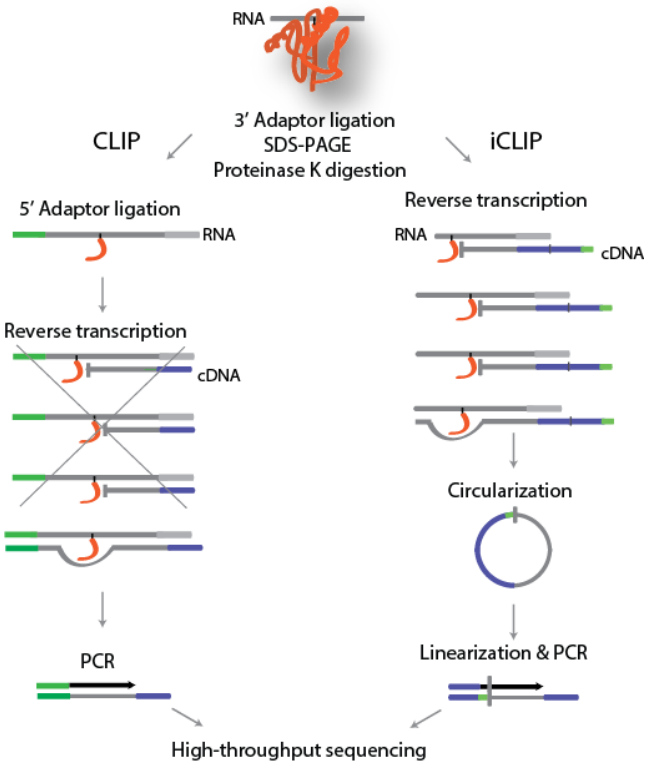


Figure 14 | CLIP and iCLIP schematic. Cartoon comparing original CLIP and iCLIP protocol. CLIP method requires two intermolecular ligations reactions, to anneal 5' and 3' adapters to the RNA. In addition, only cDNAs that contain the 5' adapter sequence are amplified in downstream PCR steps; those cDNA which do not contain the 5' adapter and are truncated, are lost. In iCLIP the 5' adapter sequence is ligated by circularisation after RT, therefore truncated cDNAs contain both 5' and 3' adapters and can be amplified in downstream PCR steps. Figure was originally published in Sugimoto et al. (2012).

into the cDNA during reverse transcription for further PCR amplification in the next step. In order to circumvent this problem, the iCLIP protocol incorporates the 5' adapter into the primers used for reverse transcription, and ligates the 5' adapter after cDNA synthesis, allowing truncated cDNAs to be recovered (Fig. 14). Additionally, the exact cross-linked nucleotide, or exact binding site for your RNA-binding protein of interest can be identified by sequencing these truncated cDNAs.

1.4.4 Key controls in iCLIP protocol

The iCLIP protocol is a 64 step protocol spread out over 7 – 8 days. It is therefore necessary to include several negative controls throughout to ensure a stringent purification, free of contaminating RNAs, which are detailed below.

1.4.4.1 No UV cross-link

UV irradiation covalently cross-links proteins and RNAs that are directly interacting, allowing highly stringent wash conditions without disrupting protein/RNA interactions. However, as all steps are performed on beads, high salt conditions can lead to loss of purified protein:RNA complexes by interfering with protein:protein interactions between antibody and the protein under study. Therefore, an optimal balance between, stringency and efficiency of the IP is required. In order to test IP stringency and that uncross-linked RNAs are not pulled down indiscriminately with the protein under study, a no UV-crosslinking control is also performed in parallel using low RNase. As only RNAs covalently cross-linked to proteins should remain after stringent IP conditions, this control lane should be devoid of signal.

1.4.4.2 High RNase sample

An attractive factor, common to all the CLIP techniques, is the ability to identify the RNA binding profile of an endogenous protein. However, specific and robust endogenous antibodies are not always available. In order to circumvent this, epitope tagged proteins are now widely used. Nevertheless, in order to avoid co-purifying other RBPs along with the protein under study, the specificity of the antibodies and the purification protocol must be assessed. Within the iCLIP protocol this can be tested at two steps, by testing a 10 % sample of the IP by western blot using specific antibodies to check for efficient IP and by lack of signal in the negative controls, and secondly, by the migration of the protein:RNA complexes isolated from samples degraded with high RNase concentrations. When separated by SDS PAGE these protein:RNA complexes should uniformly migrate at a similar point of that of uncross-linked proteins. In addition, if other bands are present it may indicate contamination with other RBP proteins co-purifying with the protein under study. Although most protein:protein interactions are disrupted during 15 high (1M salt) and low salt washing steps, particularly strong protein:protein interacts can remain, and therefore the IP protocol must be modified accordingly before proceeding further.

1.4.4.3 Processing of negative IP RNAs in parallel

In addition, it is important to process protein:RNA complexes isolated from a negative control IP from samples partial degraded with a low RNase control. Generally, negative control IP antibodies are IgGs matching the species of that antibody used for isolating the endogenous protein of interest. If using epitope tagged proteins, cells are transfected with an empty vector and beads against the tag chosen are used for negative control IPs. Isolating RNAs from these negative controls in parallel with the protein under study allows the assessment of the stringency of the protocol. At the penultimate step, Illumina barcodes adapters are added to the cDNA libraries by PCR. Here DNA-PAGE is used to assess library quality. Firstly, no amplification should be observed in the negative control other than primer dimers which run at 141nt. Amplification indicates contaminating nucleic acids that have been carried over during the protocol. Finally, if the ratio of signal for primer dimers is higher than that of the amplified library, this could indicate low library complexity, which is caused by low amounts of input material.

1.5 Part 5: Outline of this thesis

There is an overwhelming body of evidence indicating the Microprocessor complex can perform important function outside of miRNA biogenesis. The identification of the direct RNA target of the RNA binding component of the Microprocessor complex, DGCR8, unveiled an expanse of previously unidentified substrates. In addition to the canonical targets of DGCR8 (pri-miRNAs), Macias et. al. (2012) identified enrichment of DGCR8 CLIP tags within mRNAs, lncRNAs and repetitive elements. Subsequently, for these RNAs it was found that the Microprocessor complex negatively regulated these RNAs directly by cleavage, leading to transcript destabilisation. However, DGCR8 was also found to bind snoRNAs reproducibly. Further investigation identified the abundance of mature snoRNAs was directly regulated in a DGCR8-dependent, Drosha-independent manner, indicating that DGCR8 could form additional complexes with other ribonucleases. These observations form the basis of my Ph.D. and the primary goal of this thesis is to characterise the alternative complex formed by DGCR8 to regulate the mature levels of snoRNAs, which has ultimately led to investigation of the exosome complex and its RNA substrates.

Chapter 2. Material and Methods

2.1 Wet Laboratory

2.1.1 Cell culture, transfections and antibodies

2.1.1.1 Cell culture

HEK293T, HeLa and SH-SY5Y cells were grown under standard conditions in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovin serum and 1% penicillin-streptomycin. Mouse Embryonic Stem cells (mESC) were grown on gelatin-coated plates (Sigma) without feeders in DMEM-high glucose supplemented with 15% (v/v) fetal bovine serum (GIBCO-Invitrogen), LIF, glutamine and essential amino acids. *Dgcr8* -/- mESC were purchased from Novus Biologicals (NBA1-19349) and the parental strain (v6.5) from Thermo Scientific (MES1402). *Dicer* -/- and Dicer f/f were kindly provided by Robert Blelloch (UCSF) (Babiarz et al., 2008).

2.1.1.2 siRNA and DNA transfections

Knock-down of endogenous proteins was performed in HeLa and SH-SY5Y cells after two rounds of siRNA transfection using Dharmafect 1 solution (T-2001-03, Dharmacon). Briefly, cells were seeded in 6-well plates to 40% confluence and after 24 h were transfected using 25 nM of each siRNA pool and 10 μl of the transfection reagent. The transfection medium was replaced after 24 h and cells were grown for another 24 h. Cells were then re-transfected following the same protocol and collected 24h after the second transfection for analyses. siRNA pools were purchased from Dharmacon, Drosha (L-016996-00), DGCR8 (L-015713-00), hRRP6 (L-010904-00), hDIS3 (L-015405-01), hRRP41 (L-013760-00), ZCCHC7 (L-014804-01), RBM7 (L-017936-02), hMTR4 (E-031902-00-0005) and non-targeting siRNAs (control) (D-001810-02). Overexpression analyses were performed in either HEK293T or HeLa cells by transfecting plasmids using Lipofectamine 2000 (11668027, Invitrogen) following standard manufacturer's protocol.

2.1.1.3 Antibodies used for Western blot experiments

Antibodies used for immunoprecipitations and Western blot analyses were the following, anti-DGCR8 antibody from Abcam (ab90579) and from Santa Cruz Biotech (sc-48473), anti-Drosha antibody from Novus Biologicals (NBP1-03349), anti-fibrillarin

(ab4566), anti-hRRP6 (ab50558), anti-hRRP41 (ab137250) from Abcam. The anti-hRRP40 antibody (sc-98776) and anti-dyskerin (sc-48794) were purchased from Santa Cruz Biotech, and the anti-hDIS3 (14689-1-AP) antibody from Protein Tech. The anti-FLAG antibody was purchased from Sigma-Aldrich (F3165, mouse and F7425, rabbit), and the T7-antibody from Merck-Millipore (65922).

2.1.2 Immunoprecipitation experiments

2.1.2.1 Epitope tagged Immunoprecipitations and Mass Spectrometry analysis

HEK293T cells were transfected with T7-tagged DGCR8 and pCG backbone as a negative control. An N-terminus 1x FLAG-DGCR8 (a gift from Sonia Guil), a C-terminus FLAG-Drosha (a gift from Narry Kim) and an empty FLAG control vector were also used. Cells overexpressing DGCR8 or Drosha proteins were collected 48 h after transfection, resuspended in buffer D (20mM Hepes-KOH pH 7.9, 100mM KCl, 0.2mM EDTA, 0.5mM DTT, 0.2mM PMSF, 5% glycerol) and sonicated with Bioruptor for 5 cycles (30 sec on 30 sec off) containing RNAse A or RNAse inhibitors, centrifuged for 10 mins at maximum speed, and the supernatant was added to beads overnight at 4°C. For purification of T7-DGCR8, T7 tag antibody agarose from Novagen was used (69026), for FLAG-Drosha and FLAG-DGCR8, anti-FLAG M2 affinity from Sigma (A220) was used. After immunoprecipitation, beads were washed for 5 times with buffer D (150 mM KCl). For mass spectrometry purposes, beads were washed 5 times with buffer D (200 mM KCl). Immunoprecipitated material were loaded in Tris-Glycine gels (Invitrogen) and run for 1-2 cm, so each individual sample was sent as a unique gel slice. Gel slices were stain with Colloidal Blue (LC6025, Life Technologies). The mixture of proteins was analysed by nLC-MS-MS at the 'FingerPrints' Proteomics Facility (University of Dundee). Proteins that co-purified with FLAG or pCG immunoprecipitations were used as a background for DGCR8 and Drosha purifications. True interactors were defined as those proteins with a mascot score ratio >1 LOG2 (IP mascot score/Control mascot score), and small and large subunit ribosomal proteins were removed from the final list of interactors (see appendix

for full list). For Gene Ontology analysis of the DGCR8-interacting proteins the DAVID Bioinformatic Resources 6.7 was used (Huang et al., 2009).

Table 5 | Table of oligos used within this thesis. Pmn = Puromycin; due to the adapter being RNA, the double deoxy cytosine added to the 3' end during the normal iCLIP protocol is replaced with a Puromycin conjugated to a ribonucleotide. X33 = phosphorylated; pre-phosphorylated oligoes used as the 5' end of this primer is ligated to the 3' end of the cDNA produced after reverse transcription. Lower case nucleotides = are adenylated tales or restriction enzyme sites. N = degenerative base.

Assay	Description	5'-3' Sequence
qRT-PCR	human mature U16	TGCCTGCTGTCAGTAAGCTG
		TGCTCAGTAAGAATTTTCGTCAA
	human precursor (3'end side) U16	TGCCTGCTGTCAGTAAGCTG
		GGCCTCCACGACACATCTAT
	human mature U92	GTCACCATGCCTCCCTAGAA
		ATCTGTCTGCCCCGTATCTG
	human precursor U92	CGGGACGAATTGAGTGAAAT
		AAAGAGGCAGGGCTAAAAGG
	human pri-miR-24-1	AGCTGAGGCGCTGCTTCT
		CCTCGGGCACTTACAGACA
	human 7SK	CATCCCCGATAGAGGAGGAC
		GCGCAGCTACTCGTATACCC
	mouse host pre-mRNA U16	TGGCTCGAATTCCAAGAGTT
		CAGTTGGTCAGTTGCCAAGA
	mouse mature U16	CTCTGTTCACAGCGACAGTTG
		TTCGTCAACCTTCTGAACCA
	mouse precursor U16 (3'end side)	TGGTTCAGAAGGTTGACGAA
		CCCACGACACATCTGTTTTC
	mouse host pre-mRNA U92	CCAAGTGCTGGGATTAAAGG
		TGTCCTCAGCACCCTAACAA
	mouse mature U92	CACTGGACCTCCCCAGAGTA
		AATTGTCTGCCCCGTATCTG
	mouse precursor U92 (3'end side)	CAGGGCGAATTGAGTGAAAT
		GCACAGGGCTGAAAGAAAAA
	mouse 7SK	GACATCTGTCACCCCATTGA
		GCCTCATTTGGATGTGTCTG
	human DGCR8 mRNA	GGCGGTCGGTCGGTGAGGCTTTC
		GGGGCTCTCATCTGTCTCCAT
	human RRP6 mRNA	ACCCAAGGACCACAGAACAG
		TCCAGCAAAAGCCTTGAAGT
	human DIS3 mRNA	GCTCTGCTTCGAAAACATCC
		GCCTGATCCAAAGACTCAGC
	hTR RNA	CCCTAACTGAGAAGGGCGTA

		TGCTCTAGAATGAACGGTGGA
	mouse TERC (mTR)	TCTTAGGACTCCGCTGCC
		CCCACAGCTCAGGTAAGACA
	human Drosha mRNA	CATGCCCGAACCTACACTG
		GGTCCTTTCCCACAGCCTAT
	human GAPDH mRNA primers	CAACGGATTTGGTCGTATTG
		GGAAGATGGTGATGGGATTT
	human RRP41 mRNA	TAGTTCAGCGACCTTCAGCA
		GAGGATGGCTGCTTCGAAAG
	human RBM7 mRNA	TTTCGAGCTTTTCCACCAGG
		GTGGGGCATGACTACTTCCT
	human MTR4 mRNA	TTCCGATGGATTTCTTCTGG
		CCAATTGTTGGGCTCATCTT
	human ZHCCH7 mRNA	GTCATCAGGGAGGAAGAGCA
		TCTGACCCATCTGACAGCTG
Telomere assay	mouse telomeric repeat quantification (Callicott 2006)	CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT
		GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCCT
	mouse genomic c-myc for normalisation (Callicott 2006)	GAGGGCCAAGTTGGACAGTG
		TTGCGGTTGTTGCTGATCTG
Cloning	hRRP6 cDNA amplification for cloning, containing restriction sites	ACAAGGAATTCGCGCCACCCAGTACCCGGGAGCC
		ACAGTGGCGGCCGCTCCTATCTCTGTGGCCAGTTGTACCTG
Mutagenesis	site directed mutagenesis dsRBD1 DGCR8	CTTGCGAAGAATAAAAAGAAACGAGCTACACTGG
		CCAGTGTAGCTCGTTTCTTTTTATTCTTCGCAAG
	site directed mutagenesis dsRBD2 DGCR8	GTTGGAAAGCAGTTAAAGAAACAGAAGATCCTTC
		GAAGGATCTTCTGTTTCTTTAACTGCTTTCCAAC
	hRRP6 mutagenesis residue D313N	GAATTGTCAGGAATTTGCAGTTAACTTGGAGCACCAC
		GTGGTGCTCCAAGTTAACTGCAAATTCCTGACAATTC
iCLIP	cut_oligo	GTTCAGGATCCACGACGCTCTTCaaaa
	Rclip 1 (IgG)	X33NNAACCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
	Rclip 2 (ENDO)	X33NNACAANNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
	Rclip 3 (FLAG)	X33NNATTGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
	R5clip (CAT)	X33NNCGCCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
	R6clip (WT)	X33NNCCGGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
	3' RNA adapter	X33UGAGAUCGGAAGAGCGGUUCAGPmn
	P5Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
	P3Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC

2.1.2.2 Endogenous Immunoprecipitations

For the purification of endogenous DGCR8 and hRRP6, 1 μg of antibody was coupled to Protein G Dynabeads (10001D, Invitrogen), in the presence of total cell extracts prepared in IP buffer (50mM Tris pH 7.5, 150 mM NaCl, 1mM EDTA and 1% Triton X100). After overnight binding, beads were washed 5 times, 10 minutes each at room temperature, with IP buffer. For RNA co-immunoprecipitations, RNA was extracted from the beads, as well as from inputs, using Trizol LS (Invitrogen) and following manufacturer's instructions. RNA samples were treated with RQ1 DNAse (Promega, M610A) for 1 h at 37°C, phenol/chloroform extracted and ethanol precipitated. For qRT-PCR experiments, input and immunoprecipitated RNA were quantified using SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit (Invitrogen, 11736-051). The immunoprecipitated RNA was normalised to the input fraction and was expressed relative to the negative control (IgG) (set arbitrarily to 1). Primers used for these analyses are listed in Table 5. For the analysis of native DGCR8 and Drosha complexes, immunoprecipitates from FLAG-DGCR8 and FLAG-Drosha were washed at low salt conditions (150mM KCl-buffer D) and eluted by using 1x and 3x FLAG peptide (F3290 and F4799, respectively) during 6 h rotating at 4°C. Beads were then centrifuged at 8000 g for 1 min, and the supernatant was used to load gradients. V5-tagged proteins were immunoprecipitated with V5-antibody coated beads (Sigma-Aldrich, A7345), following the same protocol as for endogenous protein immunoprecipitations (IP buffer).

2.1.3 Plasmid and mutant construction

FLAG-hRRP6 was amplified from a cDNA clone (Origene) using the primers listed in Table 5 and cloned in a pcDNA3-3xFLAG vector. The FLAG-hRRP6 catalytically inactive mutant (D313N) was generated by Quickchange site-directed mutagenesis kit from Stratagene by mutating the wild-type sequence (FLAG-hRRP6 plasmid), using oligos listed in Table 5, and was based on the mutation described previously (Januszyk et al., 2011). The mutant DGCR8 expression vector (T7-DGCR8 dsRBD1&2 mut) was generated by Quickchange site-directed mutagenesis kit from Stratagene (200521) based on Yeom et al., (2006). For all the primers used for cloning see

Table 5. The truncated DGCR8 (1-726) was amplified from FLAG-DGCR8 plasmid using primers designed to create a C terminal truncation at amino acid residue 726 (Yeom et al., 2006), and then cloned into pcDNA3-3xFLAG vector. All mutants sequence integrity was confirmed by in house sequencing services.

2.1.4 Quantitative and semi-quantitative RT-PCR analyses

2.1.4.1 RNA extraction and quantitative RT-PCR

Total RNA was isolated from cells using Trizol (Invitrogen) and treated with DNAse (RQ1 DNAse, Promega, M601A) and checked for DNA contamination by PCR. For one step qRT-PCR, 500-100 ng of total RNA was used with SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit (Invitrogen, 11736-051) on CFX96 real time system. Data was analysed with Bio-Rad CFX Manager software. All experiments, unless stated, show the average and standard error of the mean of at least three independent biological replicates. Primers for qRT-PCR analysis are listed in Table 5.

2.1.4.2 Telomere length analysis and DNA extraction

Average telomere length was measured by real time quantitative PCR from genomic DNA extracted from *Dgcr8* +/+, *Dgcr8* -/-, *Dicer* +/+ and *Dicer* -/- cell lines following a previously described protocol (Callicott and Womack, 2006). Primers used for this experiment are listed in Table 5.

2.1.5 Subcellular localisation and native complex analyses

2.1.5.1 Immunofluorescence

Cells were grown on coverslips, fixed with 4% paraformaldehyde, washed with PBS and permeabilised with 0.2% Triton X-100. Coverslips were then incubated for 1 hour with blocking buffer (1% BSA, 0.01% Triton X-100 in PBS). Primary antibodies used to detect nucleolin (MA1-20800, Thermo Scientific), FLAG (F7425, rabbit), hRRP6 (ab50558, Abcam), Drosha (NBP1-03349, Novus Biologicals) and T7 overexpressed proteins (Merck-Millipore, 65922); and secondary antibodies used were anti-rabbit Alexa Flour® 488 (A11070, Molecular Probes) and anti-mouse Alexa Flour® 594 (A21203, Molecular Probes). All antibodies were diluted to working concentrations in diluted blocking buffer (1:1000) and incubated with coverslips in humidified chamber overnight

at 4°C. Coverslips were washed 3 times with wash buffer (0.01% Triton X-100 in PBS). Secondary antibodies were incubated with coverslips in a dark, humidified chamber for 2 hours at room temperature. Coverslips were then washed with washing buffer and mounted with DAPI containing mountant (P-36931, Thermo Scientific). Images taken on Ziess Axioplan2 with an objective lens mounted to a PIFCO collar. Images were deconvolved and processed using Volocity.

2.1.5.2 Nucleolar purification

Ten 150mm plates of HeLa cells were collected by trypsinisation when ~90% confluent. Nucleolar purification was essentially carried out as previously described (http://www.lamondlab.com/f7nucleolarprotocol.htm) with one modification; S3 buffer was supplemented with 0.1mM $MgCl_2$. Antibodies employed as nucleoplasmic markers were anti-Lamin B (ab16048, Abcam) and anti-eIF4A3 (ab115022, Abcam); and as a nucleolar marker anti-fibrillarin (ab4566-250, Abcam).

2.1.5.3 Glycerol gradients

Glycerol gradients (5-30%) were poured using the Biocomp gradient station model 153 (BioComp Instruments, Inc. New Brunswick, Canada) and contained 50mM Tris pH .5, 150mM NaCl, 1mM EDTA. Nuclear HEK293T extracts or eluted FLAG immunoprecipitates were loaded in 11ml 5-30% glycerol gradients and centrifuged for 16 h at 41,000 rpms in a Sorvall SW41Ti rotor. Following centrifugation, fractions (500 µl) were collected manually from the top. When just showing heavy and light fractions from the gradient, the first 11 fractions were pooled and precipitated with TCA and the same applies for the next 11 fractions. All fractions were precipitated using standard TCA precipitation, the pellet was resuspended in loading buffer and analyzed in Tris-Glycine 4-12% gels.

2.1.6 iCLIP protocol

The protocol outlined below is based on Huppertz et al. (2013), with a few modifications, such as, an 3'RNA linker is used instead of a DNA linker and therefore the ligation steps have been altered appropriately. HEK293T cells were expanded in 15cm dishes to 95% confluencey. Next, the media was replaced with 6ml of ice cold PBS and

cells were UV-irradiated on ice with 1200J at 254 nm, then rotated by 180° and repeated. Cells were then collected by scraping and washed with PBS followed by centrifugation to pellet cells. Pellets were then frozen on dry ice and stored at -80°C until required. Cell pellets were defrosted and resuspended in 1 mL lysis buffer (50 mM Tris-HCl, pH 7.4; 100 mM NaCl; 1 mM $MgCl_2$; 0.1 mM $CaCl_2$: 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate; 0.1% SDS; 1x Protease inhibitor [04693132001, Roche], 1/1000 anti-RNase [AM2692; Life Technologies]). Cell lysates were generated by sonication for 5 cycles of 30 seconds on/off on ice followed by centrifugation at full speed to pellet cell debris and the supernatant was retained as the cell lysate. Dilutions of RNase I (Ambion, AM2295) 1/500 (low RNase) and 1/15 (high RNase) were made. Next, cell lysates were supplemented with 10 µl of low or high RNase I dilution as well as 2 µl of Turbo DNase (AM2238, Life Technologies) and incubated for exactly 3min at 37°C, shaking at 1100 rpm. Each immunoprecipitation was carried out in three conditions; no-UV, low RNase and high RNase, to control for immunoprecipitation specificity and also non-specific labelling Immediately following this, samples were incubated on ice for 3 mins, and then span at full speed for 20 mins at 4°C to pellet debris and supernatant was retained, leaving roughly 50 µl of lysate with the pellet. Cell lysates were pre-cleared by incubating them with 10 µl of Protein A Dynabeads (10001D) for 1 hr at 4°C, and then span at max speed to pellet beads. Again, cell lysates were retained, leaving roughly 50 µl of lysate with the pellet.

Rabbit IgG (10 µg) or anti-hRRP6 (10 µg) (ab50558, abcam) were pre-incubated with 100 µl Protein A Dynabeads for 1 hr, rotating at room temperature. Following this Protein A Dynabeads beads or M2 FLAG Dynabeads (40µl) (M8823, Sigma-Aldrich), were washed 3 time with lysis buffer, and left in final wash until ready to proceed.

Next, wash buffer was removed from the beads and cleared cell lysates were incubated with the beads overnight, rotating at 4°C. The next day, immunoprecipitates were washed 3 times with high salt buffer (50 mM Tris-HCl, pH 7.4; 1 M NaCl; 1 mM EDTA; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate) and two times with PNK buffer (20 mM Tris-HCl, pH 7.4; 10 mM MgCl2; 0.2% Tween-20). All wash steps were

carried out on rotating wheel for 5 mins. The following steps are performed 'on-bead' unless stated otherwise. Next, 3' ends of RNA from RNA:Protein complexes were dephosphylated by incubating beads in 20 µl of CIP mix (15 µl water; 3 µl 10x CIP buffer; 1 µl CIP enzyme [M0290S, NEB]; 1 µl RNasin [N2515, Promega]), and incubated for 30 mins at 37°C shaking 1000 rpm. Beads were then washed as follows: 1 x PNK buffer 2 x high salt buffer and 2 x PNK buffer. Next, 3' RNA linker were ligated to RNA:Protein complexes isolated on bead by incubating beads for 16 hr at 16°C, shaking at 1000 rpm (every 1.5 mins for 30 seconds) in 20 µl of ligation mix (3 µl water; 3 µl 10x T4 RNA ligase buffer; 1 µl RNA ligase [M0204S, NEB]; 2µl ATP [10mM]; 1 µl RNasin; 2 µl pre-adenylated 3' RNA adapter [50 µM]; 6 µl of 50% PEG6000; 2µl DMSO). The next day samples were washed as follows: 2 x high salt buffer and 3 x PNK buffer. Following this, RNA:Protein complexes were then labelled with $_{\gamma}$32-ATP (Perkin Elmer) by incubating beads with 8 µl of labelling mix (7 µl water; 0.4 µl $_{\gamma}$32-ATP; 0.4 µl 10 x PNK buffer; 0.2 µl T4 PNK enzyme [M0201L, NEB]) for 5 mins at 37°C shaking at 1000rpm. The PNK mix was then aspirated and beads were washed once in PNK buffer. Beads were then mixed with 15 µl NuPage loading buffer mix (6 µl 4x NuPAGE LDS sample buffer[NP0007, ThermoFisher Scientific]; 3.2 µl 10 x NuPAGE sample reducing agent [NP0004, ThermoFisher Scientific]; 5.8 µl PNK buffer) and incubated at 70°C for 10 mins at 1000rpm. RNA:Protein complexes were then separated on 3-8% Tris-Acetate gels (EA0375BOX, ThermoFisher Scientific) by SDS-PAGE, and then transferred to nitrocellulose membrane (Hybond-N, RPN203N, GE healthcare). Next, membranes were exposed to X-ray films (28906836, GE Healthcare) at -80°C for 30mins, 1 hour and overnight exposures to visualise labelled RNA:Protein complexes.

X-rays were used as a guide to cut from membranes above the expected size of hRRP6 (~120kDa) in low RNase treated samples. The isolated membranes were shredded into many pieces and were incubated 210 µl of protease mix (200µl PK buffer [100mM Tris-HCl, pH 7.4; 50mM NaCl; 10mM EDTA]; 10µl proteinase K [YSJ-762-Q, Fisher Scientific]) for 20 mins at 37°C shaking at 1000 rpm. Next, 200 µl of PK/7M urea mix (0.42 g urea per 1 mL of PK buffer) was added to the membranes and incubated for a further 20 mins at 37°C shaking at 1000 rpm. Samples were then briefly centrifuges to

pellet membrane pieces. The supernatant and 400 µl of RNA phenol chlorophorm (AM9720, ThermoFisher Scientific) were added to heavy phase lock gel tube (713-2536, VWR), and incubated for 5 mins at 30°C at 1100 rpm. Following this, tubes were centrifuged at 13,000 rpm for 5 mins at room temperature. Following this, the upper phase was collected and mixed with 40 µl of NaAC (3M, pH 5.5) and 1 µl Glycoblue (9510, Ambion) and mixed; then, 1 mL of 100% ethanol was also added and left to precipitate overnight at -20°C.

The following day, precipitations were span for 45mins at 4°C at 13000 rpm. Pellets were washed with 500µl of 80% EtOH and resuspended in 6.25 µl of water in preparation for cDNA synthesis. Next, to the resuspended RNAs 0.5 µl of Rclip primer (Rclip 1-4 [0.5pmol/µl], see Table 5) and 0.5 µl dNTP (10mM) were added and then tubes were incubated for 5 mins at 70°C. Next, samples were cooled to 25°C and 2.75 µl of reverse transcription mix (RT) was added (2 µl 5x RT buffer; 0.5 µl 0.1M DTT; 0.25 µl Superscript III RT [18080085, Life Technologies]) and incubated as follows: 5 mins at 25°C, 20 mins at 42°C, 40 mins at 50°C and 5 mins at 80°C before cooling to 4°C. Next, cDNAs precipitated by mixing with 90 µl TE buffer, 1 µl Glycoblue and 10 µl NaAC (3M, pH 5.5), 250 µl 100% ethanol and incubated overnight at -20°C.

The next day precipitates were span for 45mins at 4°C at 13000 rpm. Pellets were washed with 500 µl of 80% EtOH and resuspended in 6 µl of water. Next, cDNA were separated by SDS-PAGE. 6 µl 2x TBE-urea loading buffer (LC6876, Life Technologies) was added to the samples and LMW ladder (1/30 [N3233L, NEB]) and heated to 80°C for 3 min and directly loaded onto a 6% TBE-urea gel (EC68652B, Life Technologies). The gel was ran for 40 min at 180V. Three bands were cut from the gel at sizes corresponding to 120-200 nt (high), 85-120 nt (medium) and 70-85 nt (low). These gel slices were then incubated with 400 µl TE and crushed into small pieces using a 1 ml syringe plunger and incubated at 37°C for 1 hr at 1100 rpm in order to elute cDNAs. Samples were snap frozen on dry ice for 2 mins and then incubated for a further 37°C for 1 hr at 1100 rpm. Following this, gel pieces were removed from the samples by filtering samples through Costar SpinX columns (8161, Corning Incorporated) loaded with two 1 cm glass pre-filters (Whatman, 1823010) and centrifuging for 1 min at 13000 rpm at room

temperature. cDNAs were then precipitated by mixing with 1 µl Glycoblue and 40 µl NaAC (3M, pH 5.5), then adding 250 µl 100% ethanol and incubated overnight at -20°C.

The next step is to circularise the cDNAs. Firstly, the precipitates were span for 45mins at 4°C at 13000 rpm. Pellets were washed with 500 µl of 80% EtOH and resuspended in 8 µl of ligation mix (6.5 µl water; 0.8 µl 10x CircLigase Buffer II; 0.4 µl 50 mM $MnCl_2$; 0.3 µl; Circligase II [CL9025K, Cambio]), and incubated for 1 hr at 60°C. After, 30 µl of oligo annealing mix is added (26 µl water; 3 µl Fast Digest buffer; 1 µl cut_oligo [10µM]; see Table 5), and incubated for 1 min at 95°C followed by a decrease of 1°C every 20 seconds until 25°C is reached. Next, 2 µl BamHI (FD0055, Fermentas) is incubated with the samples for 30 mins at 37°C. Samples were then transferred to 1.5 mL tubes and the now linearised cDNAs are precipitated by mixing with 50 µl TE, 0.5 µl Glycoblue and 10 µl NaAC (3M, pH 5.5), then mixed with 250 µl 100% ethanol and incubated overnight at -20°C.

Next, cDNAs are amplified by PCR and Illumina sequencing adapters are also added during this step. Firstly, the precipitates were span for 45 mins at 4°C at 13000 rpm. Pellets were washed with 500 µl of 80% EtOH and resuspended in 11 µl of water. The PCR mix is prepared as follows: 10 µl cDNA, 1 µl P5/P3 solexa primer mix (see Table 5 [10µM]), 20 µl Accuprime Supermix 1 enzyme (12342028, Life Technologies) and 9 µl water. The PCR samples were then ran on the following programme: 94°C for 2 min, (94°C for 15 sec, 65°C for 30 sec, 68°C for 30 sec, repeat 34 cycles), 68°C for 3 min, 4°C forever. An analytical samples of 4-8 µl of the PCR products was run on 6% TBE-urea gels. The PCR products were then purified using Agencourt AMPureXP beads (A63880, Beckman Coulter), and equal volumes of libraries were mixed and then multiplexed samples sequenced on one lane of Illumina HiSeq system (Beijing Genomics Institute).

2.2 Bioinformatic Analyses

2.2.1 Analysis of DGCR8 HITS-CLIP and RRP6-iCLIP libraries

Reads were pre-processed using custom python scripts. First, reads were demultiplexed (see Rclip primers in Table 5 for barcodes) according to their fixed barcode allowing up to 1 mismatch. Next, reads were trimmed to remove low quality scores and 3' adapter sequences. Finally, duplicated reads containing identical random barcodes were removed and the 5' random barcodes were trimmed. After these steps, all reads longer than 19 nucleotides were further analysed. Reads were mapped to the human genome (hg19) using bwa-pssm (Kerpedjiev et al., 2014). All reads that were not confidently mapped (posterior probability PP <= 0.99) were then mapped to an exome index containing all collapsed exons from human ensembl70 transcripts (Flicek et al., 2014). Finally, all reads unmapped to the genome were mapped to an exon-junction index containing all annotated unique exon-junctions from human ensembl70 transcripts. Only reads mapped at any of the steps with a PP > 0.99 were considered for further analysis.

Reads from the 2 CLIP biological replicates were pooled and clustered together according to their genomic positions. Significant clusters were calculated using Pyicos (Althammer et al., 2011) and only significant clusters with a false discovery rate (FDR) < 0.01 were considered for further analysis. DGCR8 significant clusters were obtained from (Macias et al., 2012) (http://regulatorygenomics.upf.edu/Data/DGCR8/). We only downloaded the data of the second replicate both for the endogenous and the overexpressed experiments. For each of the datasets, the coordinates of the significant clusters were mapped from hg18 to hg19 using liftOver tool (Kuhn et al., 2013). To define a set of reproducible DGCR8 CLIP clusters, T7-DGCR8 and endogenous DGCR8 significant clusters were overlapped according to their genomic coordinates using fjoin (Richardson, 2006). DGCR8 reproducible clusters and hRRP6 significant clusters were overlapped based on their genomic coordinates using fjoin. These common clusters were then overlapped with the ensembl70 annotation using fjoin to identify a set of target genes bound by both DGCR8 and hRRP6.

2.2.2 Mapping RRP6-iCLIP to rDNA 45S consensus sequence

Raw sequences data from the iCLIP experiments were processed as follows: PCR duplicates were collapsed using custom perl scripts. Next, Trimmomatic was used to pre-process reads and remove Illumina adapters from the 3' end (Bolger et al., 2014). Low confidence nucleotide calls were trimmed from the 3' end using the sliding window method and an average Phred score of 20 over 3nt windows and reads below 18nt were discarded (ILLUMINACLIP:../illuminaadapter.fasta:2:30:10 SLIDINGWINDOW:3:20 MINLEN:18). Reads were demultiplexed using the FASTX Toolkit by Hannon Lab (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Finally, 10nt were trimmed from the 5' end of the read to remove the degenerative and specific barcode sequence. Reads were mapped using tophat2 (Kim et al., 2013)(--max-multihits 20 --num-threads 8 --read-mismatches 2 --library-type fr-firststrand --no-novel-indels -no-novel-juncs --no-coverage-search) to a 45S (NR_046235.1) rDNA consensus sequence downloaded from NCBI.

2.2.3 Calculating read density spanning ncRNA loci

Raw reads from the RRP6 CLIP experiments were intersected with significant clusters in order to generate a data set of statistically significant reads. Next, snoRNA sequences were extracted from GENCODE Release 19 (GRCh37.p13) and converted to bed file format. Customised bed files of 50nt windows immediately upstream (5'flanking) and downstream (3'flanking) of the annotated snoRNA were generated. The location of iCLIP reads mapping to snoRNA loci reads were the classified into 5'flanking, mature or 3'flanking; using the strategy depicted in (Fig. 37 A). Reads classified as 5' flank were those reads that overlapped 5' flank by at least 1 base; while reads classified as 3' flank were cthose reads that overlapped the 3' flank by at least 1 base. These reads were then removed from the dataset, and only those reads that mapped exclusively to mature sequences, with no overlap in to the 3' or 5' flanks, were classified as mature snoRNA read. Counts were normalised to the number of mapped reads for the dataset (reads per million) and then divided by number of reads mapping to snoRNA loci (5'flank + mature + 3' flank) to generate a distribution of reads over snoRNA loci. rules applied for 3' flank

Additionally, a previously published sRNAseq (20-200nt) data set generated in HEK293T cells was also included in these analyses (GSE43666; Kishore et al., 2013). These analyses were done with bedtools and customised scripts.

2.2.4 Calculating 3' extended tRNAs and snoRNAs

The length of 3' extended ncRNAs was calculated by first collapsing all overlapping reads into contiguous genomic intervals (bedtools). Then, these were intersected (bedtools) with all annotated snoRNAs (GENCODE release 19), and distances from the 3'end of the annotated gene (3'EAG) to the 3' end of the genomic interval was calculated using customised scripts.

2.2.5 Average read density profiles

To make standardised profiles, annotated annotated SNORD, SCARNA and SNORA with at least 30% RNA-seq coverage in HEK293 cells were selected from ensembl70 annotation. The RNA-seq used to calculate the RNA-seq coverage of the snoRNAs was a pool of published RNA-seq datasets from (Baltz et al., 2012; Kishore et al., 2011) (GSM714684; GSM714685; GSM940576), which were mapped using the same pipeline described above.

To make the profiles, each annotated snoRNAs was divided in 20 equally sized bins. We also included 50nt at each flank of the annotated snoRNA, which were divided in 5 equally sized bins. In each bin we calculated the mean enrichment per nucleotide of iCLIP $\bar{e}_b$ for bin b as:

$$\bar{e}_b = \frac{1}{g}\sum_{i=1}^{g}\frac{1}{l_i}\sum_{j=1}^{l_i}\frac{e_i}{10^6}$$

Where g is the total number of snoRNAs considered, l_i is the length of the bin and e_i is the number of iCLIP reads mapped to the bin. The same process was followed to generate standardised profiles for histone mRNAs and tRNAs.

Chapter 3. The Exosome interacts with DGCR8 but not Drosha

3.1 INTRODUCTION

In this chapter we aim to identify the putative ribonuclease/s that interact with DGCR8 and function to regulate mature snoRNA levels. To do so, we characterised the DGCR8 and Drosha interactome by performing parallel immunoprecipitations coupled with mass spectrometry analyses (IP-MS) and generated interaction profiles. Importantly, as the DGCR8-mediated regulation of snoRNA does not require Drosha; we identified proteins that specifically interact exclusively with DGCR8. Within this pool of proteins we found subunits of the exosome including the exosome-associated 3'→5' exoribonuclease RRP6.

3.2 RESULTS

3.2.1 Immunoprecipitation coupled with mass spectrometry analysis identified the DGCR8 and Drosha interactome.

The DGCR8-dependent regulation of mature snoRNAs characterised by Macias et al. (2012) was found to be Drosha-independent(Macias et al., 2012)(Macias et al., 2012)(Macias et al., 2012)(Macias et al., 2012). This raised the interesting hypothesis that DGCR8 can form additional complexes with alternative ribonucleases, other than Drosha. In order to identify this putative ribonuclease we decided to characterise the DGCR8-interactome by IP-MS analyses of two different tagged versions of DGCR8 (T7 and FLAG) in parallel with Drosha (FLAG) and their respective controls. Purifications were performed in the presence of DNase, however, RNase was omitted in order to recover RNA-dependent interactions. Whole lanes were sent for mass spectrometry analysis and a representative gel depicting 10% of IPs is shown in Figure 15 A. In order to define the interacting partners for DGCR8 and Drosha, we calculated the ratio of the mascot scores (IP mascot score/Control mascot score) and selected for proteins with a LOG2 ratio ≥1 to reduce nonspecific interactions. After removing ribosomal proteins, which are highly abundant within the cell (RPL & RPS proteins; 74 removed in total), we identified a total of 117 interactors with FLAG-DGCR8; 224 interactors with T7-DGCR8 and 80 interactors with FLAG-Drosha (Fig. 15 B; see appendix for full list of gene names).

3.2.2 Microprocessor-associated proteins

Intersection of the filtered interaction profiles identified 38 proteins that interacted with FLAG-Drosha and at least one epiptope tagged-DGCR8 IP-MS experiment (FLAG- and/or T7-) (Fig. 15 B see “Microprocessor-associated”). From these analyses, 11 have been previously identified as Microprocessor-associated proteins (Nucleolin, SRPK1, TARDBP, DDX17, DDX5, DDX15, HNRNPU, HNRNPUL1 and HNRNPA1) and, as expected the two integral components of the minimal Microprocessor, DGCR8 and DROSHA were also identified (Gregory et al., 2004; Shiohama et al., 2007) (for full list see appendix). Moreover, we identified FUS, HNRNPL and KHSRP, within

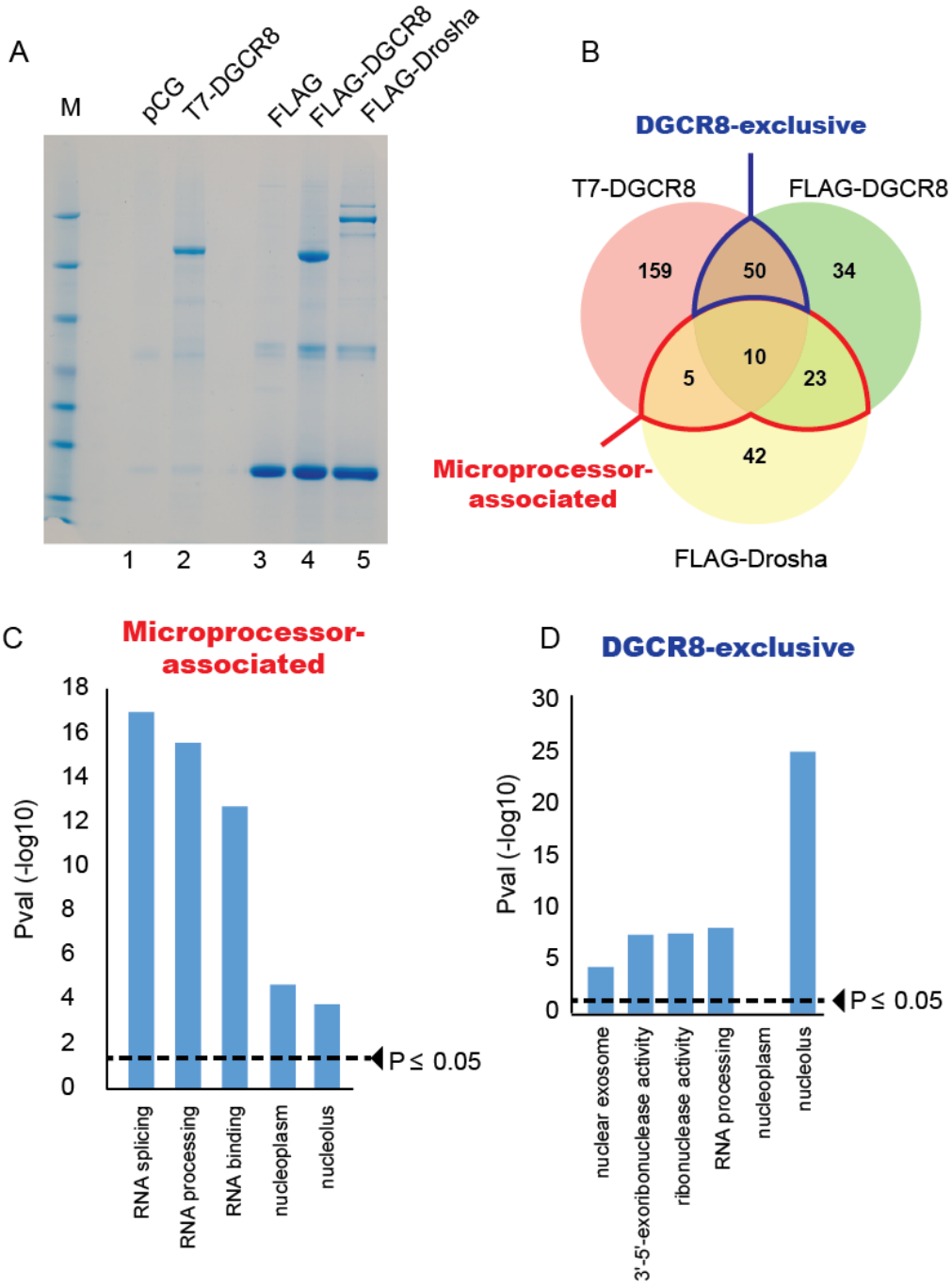


Figure 15 | Interaction profiling of the Microprocessor components, DGCR8 and Drosha. (A)Representative protein gel immunoprecipitations of pCG (T7-control), T7-DGCR8, FLAG (FLAG control), FLAG-DGCR8 and FLAG-Drosha that were sent for Mass spectrometry (MS) analyses. (B) Representation of the number of interacting partners identified by of immunoprecipitations proteins seen in (A). (B)Microprocessor-associated was defined as proteins that were identified in FLAG-Drosha immunoprecipitations that were also identified in T7- and/or FLAG-DGCR8 IPs. Whereas proteins identified in both MS analyses of T7- and FLAG DGCR8 immunoprecipitations were defined as DGCR8-exclusive. Gene ontology analyses of the 38- Microprocessor-associated proteins (C), or the 49 DGCR8-exclusive interacting partners (D) (horizontal dashed line represents significance, p-value ≤ 0.05). Values plotted are –log10, therefore values above dashed-line are significant. (E)DGCR8-exclusive proteins identified by IP-MS analyses.

the 42 proteins that co-purified exclusively with Drosha IP-MS analyses, which were previously described as interacting with the Microprocessor (Gregory et al., 2004; Morlando et al., 2012; Trabucchi et al., 2009). Gene ontology (GO) analyses of Microprocessor-associated proteins identified in our study highlighted a significant enrichment for RNA processing, binding and splicing GO terms (Fig. 15 C) (Huang et al., 2009). Additionally, nucleoplasmic GO terms were enriched (Fig. 15 C), which is in agreement with the subcellular localisation of co-transcriptional processing of miRNAs, which is the main function of the Microprocessor complex.

Table 6 | 50 DGCR8 exclusive proteins identified by IP-MS analyses seen in Figure 15. Components of the human Exosome complex highlighted in red.

Gene Symbol	Gene Name	Gene Symbol	Gene Name
ALYREF	THO complex subunit 4	ILF3	Interleukin enhancer-binding factor 3
BRIX1	Ribosome biogenesis protein BRX1 homolog	LUC7L	Putative RNA-binding protein Luc7-like 1
C7orf50	Uncharacterized protein C7orf50	LYAR	Cell growth-regulating nucleolar protein
CDC5L	Cell division cycle 5-like protein	MINA	Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA
CMSS1	Protein CMSS1	MMTAG2	Multiple myeloma tumor-associated protein 2
DDX21	Nucleolar RNA helicase 2	MYBBP1A	Myb-binding protein 1A
DHX30	Putative ATP-dependent RNA helicase DHX30	NIFK	MKI67 FHA domain-interacting nucleolar phosphoprotein
DIMT1	Probable dimethyladenosine transferase	NOP2	Putative ribosomal RNA methyltransferase NOP2
EBNA1BP2	Probable rRNA-processing protein EBP2	NPM1	Nucleophosmin
EIF2S3	Eukaryotic translation initiation factor 2 subunit 3	PURA	Transcriptional activator protein Pur-alpha
EIF6	Eukaryotic translation initiation factor 6	RBM28	RNA-binding protein 28
EXOSC10	Exosome component 10 RRP6	RBM34	RNA-binding protein 34
EXOSC2	Exosome complex component RRP4	RPF2	Ribosome production factor 2 homolog
EXOSC3	Exosome complex component RRP40	RRP1	Ribosomal RNA processing protein 1 homolog A
EXOSC4	Exosome complex component RRP41	RSBN1L	Round spermatid basic protein 1-like protein
EXOSC6	Exosome complex component MTR3	RSL1D1	Ribosomal L1 domain-containing protein 1
EXOSC9	Exosome complex component RRP45	SDAD1	Protein SDA1 homolog
GNL3	Guanine nucleotide-binding protein-like 3	SRP14	Signal recognition particle 14 kDa protein
GRWD1	Glutamate-rich WD repeat-containing protein 1	SRP72	Signal recognition particle subunit SRP72
GTPBP4	Nucleolar GTP-binding protein 1	SRP9	Signal recognition particle 9 kDa protein
H1FX	Histone H1x	SSB	Lupus La protein
H3F3C	Histone H3.3C	STAU1	Double-stranded RNA-binding protein Staufen homolog 1
HIST1H1C	Histone H1.2	TRMT1L	TRMT1-like protein
HIST1H2BJ	Histone H2B type 1-J	UPF3B	Regulator of nonsense transcripts 3B
IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	XRN2	5'-3' exoribonuclease 2

3.2.3 DGCR8-exclusive proteins

This differential proteomic approach resulted in the identification of 50 proteins that interact exclusively with DGCR8 (Fig. 15 B; see "DGCR8-exclusive"). GO analyses of these DGCR8-exclusive proteins highlighted a significant enrichment for nucleolar-associated proteins, RNA processing factors, exonuclease activity and the nuclear exosome complex (Fig. 15 D). Strikingly, IP-MS analyses revealed that 5 out 9 proteins, comprising both components of the core of the exosome complex, as well as the catalytic subunit, RRP6, interacted with DGCR8 but not with Drosha. Moreover, all exosome core proteins and RRP6 were present in T7-DGCR8 IP-MS analyses (see appendix for full list of interactors).

In addition to DGCR8-exclusive proteins we identified that T7- and FLAG-DGCR8 interacted exclusively with 159 and 34 protein, respectively (see appendix for full list of interactors). However, in order to identify true DGCR8-exclusive interactors, we decided to focus on those proteins identified in both DGCR8 IP-MS experiments (Fig. 15 B "DGCR8-exclusive"; Table 6).

3.2.4 Validation of identified interactions

The proteomics results described above were validated with immunoprecipitations of the three overexpressed proteins (T7-DGCR8, FLAG-DGCR8 and FLAG-Drosha) in the presence or absence of RNase, followed by Western blot analysis with specific antibodies. We confirmed the interaction of DGCR8 with RRP6 and with subunits of the core exosome complex, RRP40 and RRP41 in an RNA-independent manner (Fig. 16 A). Additionally, we also tested if DGCR8 could purify components of the snoRNP complex (Dyskerin and Fibrillarin), however, these interactions were diminished in the presence of RNase, suggesting these interactions are mediated through RNA (Fig. 16 A). Importantly, we found that DGCR8 did not co-purify with DIS3, an additional catalytic subunit of the nuclear exosome complex, suggesting that DGCR8 specifically associates with the nucleolar form of the exosome complex (Fig. 16 A). We also confirmed that endogenous DGCR8 could co-IP endogenous RRP6 and, accordingly, endogenous RRP6 could also co-IP endogenous DGCR8 in an RNA independent manner

(Fig. 16 B & C). Crucially, of the proteins we tested, no co-purification with Drosha was detected, except for its canonical partner within the Microprocessor, DGCR8 (Fig. 16 A). Empty vector negative controls did not immunoprecipitate any of the proteins tested (Fig. 16 A). All of which strongly suggests the complex formed by DGCR8 and the exosome does not include Drosha.

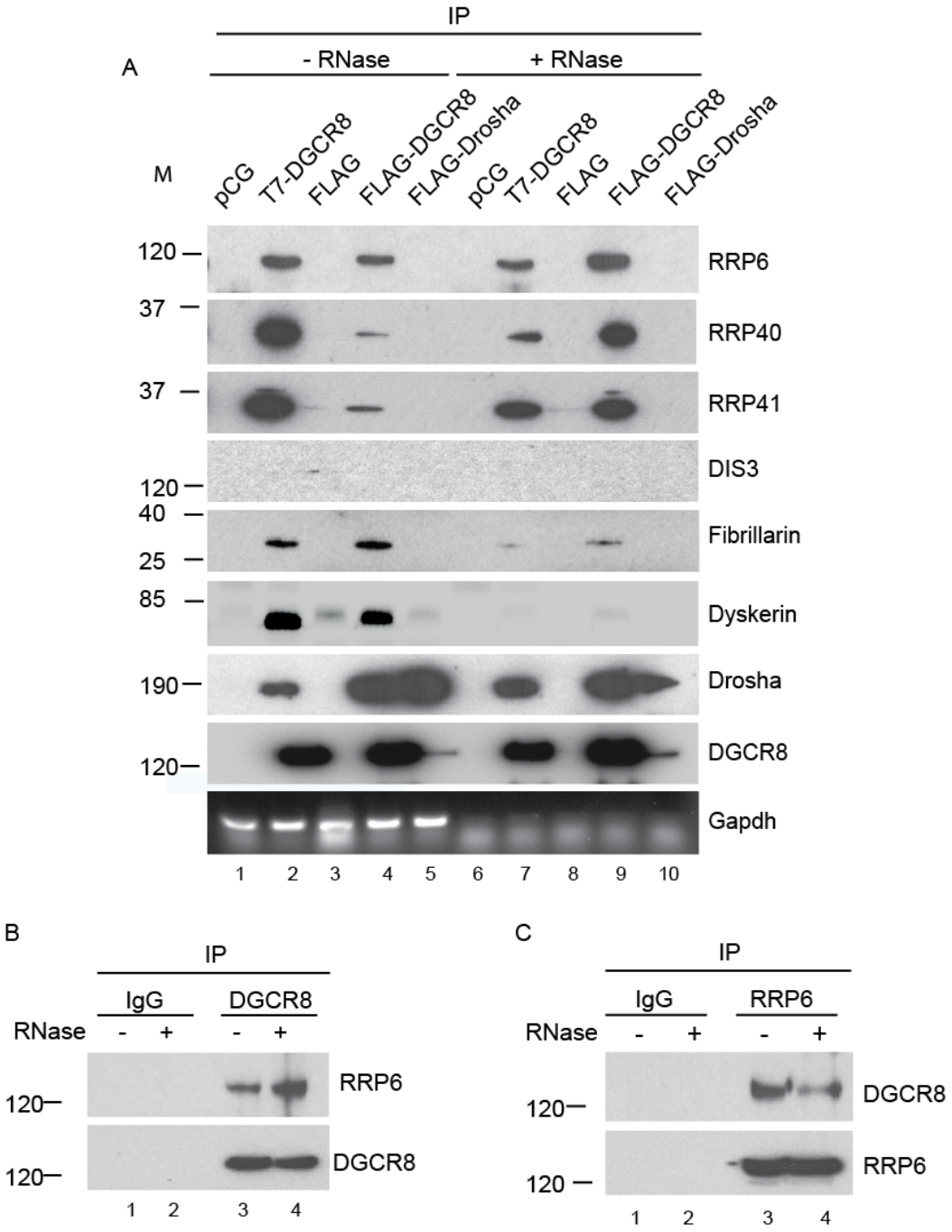


Figure 16 | Validation of DGCR8 exclusive interactors. (A) Validation of proteins interacting with T7-DGCR8, FLAG-DGCR8 and FLAG-Drosha by immunoprecipitation followed by Western blot analysis with specific antibodies, in the presence (lanes 6-10) or absence of RNase A (lanes 1-5). The RT-PCR amplification of Gapdh serves as a control for RNase treatment (bottom panel) (B, C) Reciprocal analysis of co-immunoprecipitated DGCR8 and RRP6 endogenous proteins by Western blot analysis with specific antibodies, in the presence (lanes 2 and 4) or absence of RNase A (lanes 1 and 3).

3.3 DISCUSSION

3.3.1 Summary

The Drosha-independent, DGCR8-mediated regulation of mature snoRNAs reported by Macias et al., (2012) led us to search for the putative ribonuclease(s) that can conduct these functions with DGCR8. To do so, we performed parallel IP-MS analyses of DGCR8 and Drosha. Overlapping interaction profiles of each protein identified a number of factors that interact with both DGCR8 and Drosha (Microprocessor-associated) or with DGCR8 exclusively (DGCR8-exclusive).

3.3.2 Microprocessor-associated proteins

Initial experiments that identified the Microprocessor isolated two complexes of different molecular weight; one minimal catalytic form comprising Drosha and DGCR8, and a heavier complex that includes DGCR8, Drosha and 18 ancillary factors (Gregory et al., 2004). Furthermore, an additional 5 DGCR8-associated proteins were identified from DGCR8 immunoprecipitates by other studies (Shiohama et al., 2007). Of all these proteins, we could only detect 3 that were common to all interaction profiles (NCL, HNRNPUL1 and SRPK1); indicating close association with the Microprocessor. In addition, we identified 6 other proteins previously associated with the Microprocessor that were present in at least one DGCR8 IP (T7 or FLAG) and Drosha (DDX17, DDX5, DHX15, HNRNPU, HNRNPA1 and TDP-43) (Gregory et al., 2004; Shiohama et al., 2007). Finally, we noted the presence of 3 other proteins, previously identified from the Gregory et al. (2004) Drosha IP-MS analyses (FUS, HNRNPL and KHSRP). These factors were not detected in the DGCR8 mass spectrometry analyses by ourselves and Shiohama et al. (2007), this may indicate a more direct interaction with Drosha over DGCR8. Co-factors interact with the Microprocessor directly or indirectly, by binding conserved elements or structures within pri-miRNAs which can modulate processing activities of the Microprocessor. For example, KSHRP and HNRNPA1 have antagonistic roles in pri-mir-let-7a-1 processing. While both can bind the conserved elements within the terminal loop of pri-mir-let-7a-1; KSHRP promotes pri-mir-let-7a-1 processing, but, during high concentrations of HNRNPA1, KSHRP is displaced by HNRNPA1, leading to reduced mir-let-7a-1

expression. Furthermore, TDP-43 and DDX17 both stimulate Microprocessor activity (Kawahara and Mieda-Sato, 2012; Mori et al., 2014), whereas FUS facilitates co-transcriptional recruitment of Drosha (Morlando et al., 2012).

Like intronic pri-miRNA processing, splicing also occurs co-transcriptionally, therefore the nascent pre-mRNA is decorated with splicing and miRNA processing machinery during transcription elongation (Ballarino et al., 2005; Morlando et al., 2008). The spliceosome and Microprocessor have also been reported to interact (Kataoka et al., 2009), and, intriguingly, these two complexes also compete for position over intro-exon junctions which encode pri-miRNAs (Melamed et al., 2013). Given that IP-MS experiments were performed in the absence of RNase, this enabled us to also recover RNA-dependent interactions. As such, Microprocessor-associated proteins identified from IP-MS analyses (ALL, FLAG-D8/F-Drosha & T7-DGCR8/F-Drosha) displayed an enrichment of terms associated with splicing (Fig. 15 C). These data emphasise the spatial localisation of the Microprocessor, in relation to other complexes associated with nascent RNA transcripts.

In sum, the Microprocessor-associated proteins identified within our IP-MS experiments display an extensive overlap with previously published Drosha and DGCR8 interactors; thus, we used this robust approach to identify DGCR8-exclusive proteins.

3.3.3 DGCR8-exclusive proteins

DGCR8 and Drosha are both present within the same complex and function at the nuclear phase of miRNA biogenesis, however, when analysing their interactors, their respective GO term profiles differed remarkably in terms of localisation and function (Fig. 15 C & D). This suggested these two proteins may contribute to other protein complexes and/or cellular processes. Conducting parallel IP-MS analyses of the Microprocessor complex components, DGCR8 and Drosha, enabled us to parse out DGCR8-exclusive interactors. A large number of nucleolar-associated proteins were detected with DGCR8 but not Drosha (Fig. 15 C & D), suggesting that DGCR8 can localise to the nucleolus, which has been previously reported (Shiohama et al., 2007).

In addition, we observed an enrichment of GO terms associated with ribonuclease activity, specifically 3'-5' exonuclease. This was indeed confirmed by the presence of 5 out 9 of the exosome core components, and the 3'-5' exonuclease RRP6 in both DGCR8 purifications. Importantly, the catalytic subunits of the eukaryotic exosome, display differential subcellular localisation. Whereas RRP6 is present throughout the cell but strongly concentrated in the nucleolus, the exonuclease DIS3 is excluded from this compartment (Tomecki et al., 2010).The fact that we did not detect an interaction between DGCR8 and DIS3, strongly suggests that DGCR8 forms a specific complex with the nucleolar form of the exosome. This observation is also supported by the specific association of DGCR8 to the core components of the snoRNP particle (Fibrillarin and Dyskerin), which are particularly highly enriched in the nucleolus through their association to snoRNAs. Critically however, we found that the interaction between these snoRNP factors and DGCR8 was sensitive to RNase treatment, which suggests that DGCR8 is present in the snoRNP through an RNA-mediated interaction.

Altogether, these data suggest that DGCR8 can form two complexes that have different subcellular localisation. Within the nucleoplasm, DGCR8 and Drosha, constitutes the Microprocessor complex to cleave pri-miRNAs and other cellular RNAs. Alternatively, within the nucleolus, DGCR8 interacts with exosome form that contains the exonuclease RRP6.

Chapter 4. Biochemical Characterisation of the DGCR8/RRP6 complex

4.1 INTRODUCTION

In the previous chapter we identified a number of Microprocessor-associated and DGCR8 exclusive proteins by mass spectrometry experiments. Of particular interest was the finding that DGCR8, but not Drosha, interacts with the RRP6-associated exosome complex. In this chapter we study in more detail this interaction with the purpose of understanding the mechanism by which the DGCR8-exosome interaction is important to control snoRNA levels in a Drosha-independent manner. There are four key points that will be addressed within this chapter:

(i) What are the biochemical and functional differences between the DGCR8/Drosha and DGCR8/RRP6 complex.

(ii) What are the mechanisms that control the formation of the DGCR8/RRP6 complex?

(iii) What are the roles of RRP6 and DGCR8 as factors that regulate the abundance of mature forms of snoRNAs?

(iv) What is the function of DGCR8 within this novel complex?

To answer these questions we have used biochemical and standard immunofluorescence techniques all of which are detailed within the material and methods section.

4.2 RESULTS

4.2.1 The Microprocessor components, DGCR8 and Drosha, have non-overlapping sedimentation profiles

To determine if DGCR8 and Drosha form additional complexes we performed sedimentation analyses of native DGCR8 and Drosha complexes. Overexpressed FLAG-tagged DGCR8 and Drosha were immunoprecipitated from HEK293T cells in the presence of RNase and DNase and eluted under native conditions to be loaded on a 5%-30% glycerol gradient. After centrifugation, fractions were collected manually from the top down and samples were analysed by Western blot using specific antibodies. Interestingly we observed that the sedimentation patterns of Drosha and DGCR8 were different. Drosha native complexes sediment from fractions 6-15 (Fig. 17 A top panel), whereas, DGCR8 native complexes sediment from fraction 6 throughout the gradient (Fig. 17 A bottom panel). In order to study the co-sedimentation of RRP6, we pooled light and heavy fractions together (as denoted within Fig. 17 A) and analysed these by Western blot. Again, we confirmed that Drosha native complexes did not contain RRP6 in either the heavy or light fractions (Fig. 17 B top panel), in agreement with (Fig. 15 B & Fig. 16 A). Conversely, we detected RRP6 co-sedimenting with both the DGCR8 heavy and light molecular weight fractions (Fig. 17 B bottom panel), suggesting DGCR8 and RRP6 can form heavy molecular weight complexes.

We next wanted to compare these results with the sedimentation profiles of endogenous DGCR8, Drosha, RRP6 and RRP41 from HEK293T nuclear lysates. First, we observed Drosha sedimenting in fractions 6-11, whereas RRP6 and RRP41 co-sedimented in fractions 12-20 (Fig. 17 C). Again, we observed that DGCR8 spanned across the gradient, in fractions 6 to 12, overlapping with Drosha profile, but also in heavier molecular weight fractions, overlapping with the exosome components in fractions 12 to 20 (Fig. 17 C). We also concluded that DGCR8 was present in the heavier molecular weight fractions whereas Drosha was not; this is in agreement with sedimentation profiles produced by native-complexes presented in Figure 17 A.

4.2.2 RRP6 does not require Drosha interacting region of DGCR8.

The fact that: (i) DGCR8-mediated regulation of snoRNAs is independent of Drosha; (ii) RRP6 is not present in Drosha immunoprecipitates; and (iii) RRP6 and DGCR8 can form a heavier molecular weight complex that does not contain Drosha,

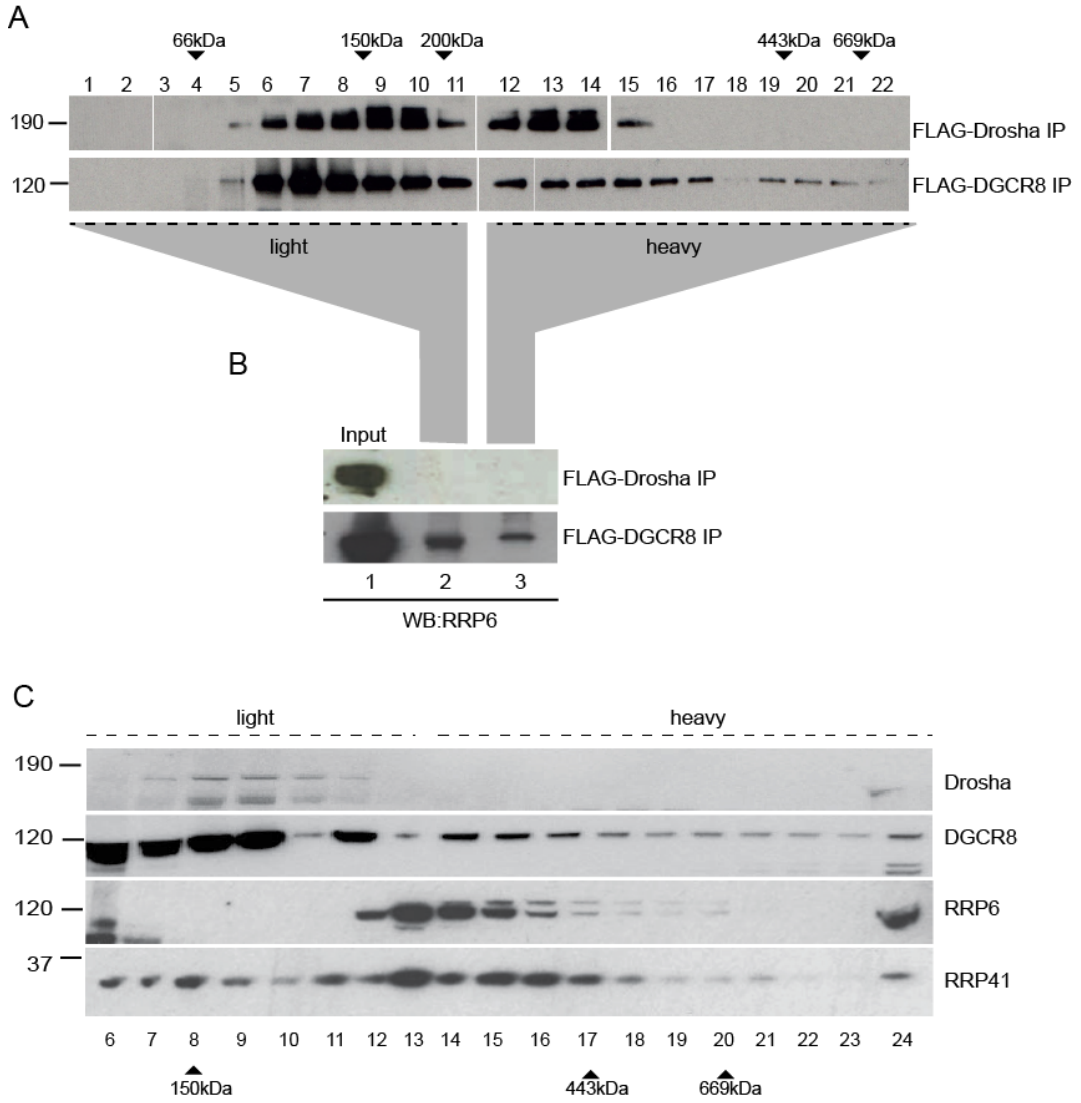


Figure 17 | Sedimentation analysis of Microprocessor and exosome components. (A) Sedimentation patterns of immunopurified FLAG-Drosha and FLAG-DGCR8 native complexes in 5-30% glycerol gradient fractions, as revealed by Western blot analysis with an anti-FLAG antibody. 'Light' denotes lighter molecular weight fractions, whereas 'heavy' indicates heavier molecular fractions. The migration of the molecular weight markers is indicated at the top (B) Western-blot of coimmunoprecipitated RRP6 with FLAG-Drosha (top panel) and FLAG-DGCR8 (bottom panel) after glycerol gradient fractionation. Fractions from a 5-30% glycerol gradient were pooled into 'light' (lane 2), corresponding to fractions 1 to 11, and 'heavy' (lane 3), corresponding to fractions 12 to 22, and run in a single lane for sensitivity purposes. (C) Sedimentation patterns of endogenous Drosha, DGCR8, RRP6 and RRP41 proteins in 5-30% glycerol gradients from nuclear HEK293T cell extracts, as revealed by Western blot analysis with specific antibodies. Lysates run in all gradients were produced in the presence of DNase and RNase.

prompted us to test if these interactions are mutually exclusive. Our hypothesis was that both Drosha and RRP6 require the same docking region of DGCR8, thus leading to steric inhibition and the formation of mutually exclusive complexes. To test this, we employed a previously characterised DGCR8 mutant lacking its C-terminus region (DGCR8 1-726) (Fig. 18 A), which was demonstrated to be essential for Drosha interaction (Yeom et al., 2006), and tested whether this domain was also required to mediate DGCR8 interaction with RRP6. We overexpressed FLAG-DGCR8 full length (FL) and a DGCR8 1-726 mutant in HEK293T cells and performed immunoprecipitations in the presence of RNase and DNase to be analysed by western blot and specific antibodies. Accordingly, FLAG-DGCR8 FL efficiently co-purified Drosha, whereas FLAG-DGCR8 1-726 could not, confirming previous studies by Yeom et al.(2006) (Fig. 18 B). Interestingly, RRP6 was efficiently co-purified with both DGCR8 versions, indicating that RRP6 does not require the Drosha-interacting region to associate with DGCR8 (Fig. 18 B). In addition we confirmed deletion of these C-terminal residues in DGCR8 did not affect its subcellular localisation (Fig. 18 C).

4.2.3 DGCR8 forms a nucleolar complex with RRP6

Results described above indicate that DGCR8 uses different domains to interact with Drosha and with RRP6, indicating that the mutual exclusivity of these complexes is not down to steric inhibition of a shared binding site. To readdress this, we hypothesised that the formation of these two DGCR8-ribonuclease complexes may be due to mutually exclusive subcellular localisation of their respective ribonucleases. It was previously suggested that the Drosha containing Microprocessor complex is present within the nucleoplasm, processing pri-miRNAs to pre-miRNAs whereas RRP6 is mainly enriched within the nucleoli (Tomecki et al., 2010). To validate these localisation patterns, we performed immunofluorescence experiments in HeLa cells using specific antibodies for endogenous Drosha and RRP6, and also Nucleolin as a nucleolar marker. We observed Drosha was restricted to the nucleoplasm and excluded from the nucleolus, whereas RRP6 was enriched within the nucleoli and also present, to a lesser degree, within the

nucleoplasm (Fig. 19 A). We also studied the subcellular localisation of overexpressed DGCR8 (T7 and FLAG-tagged DGCR8), since there is a current lack of good quality

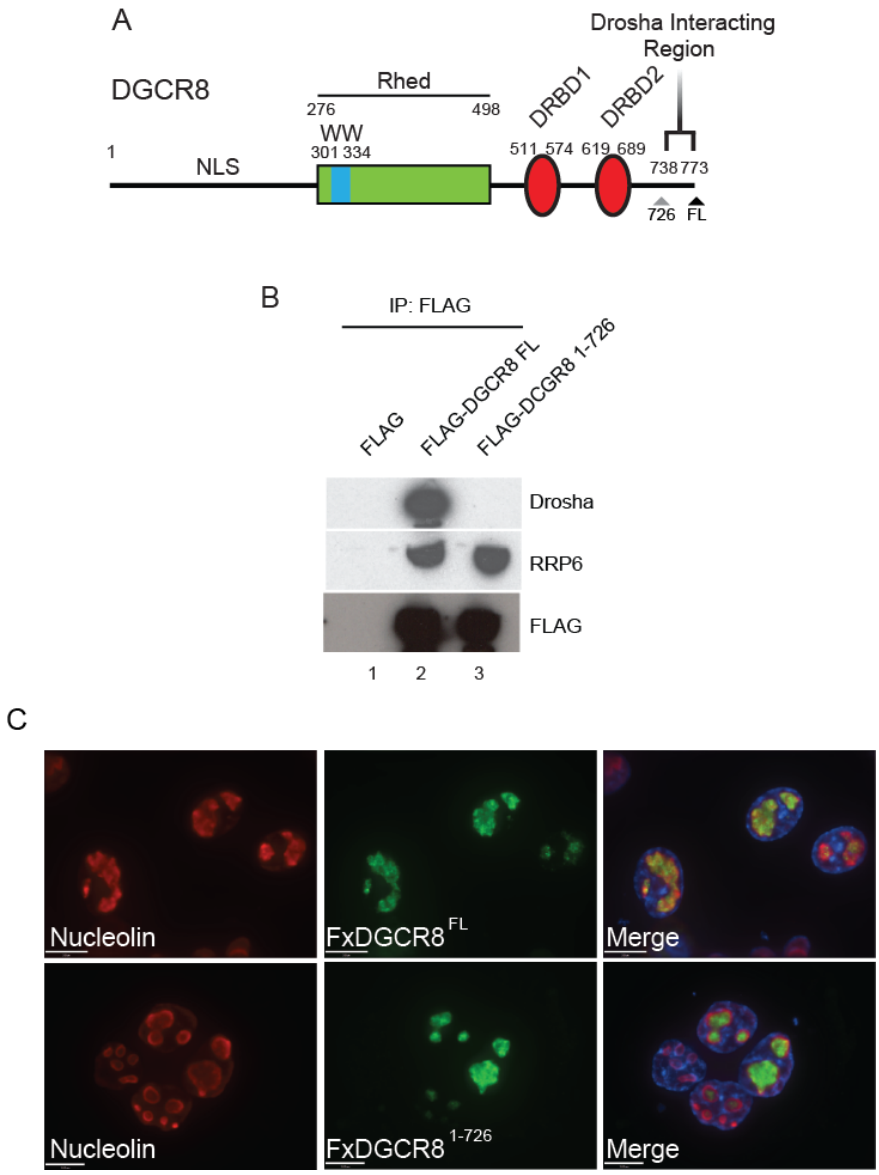


Figure 18 | RRP6 and Drosha do not interact with the same region of DGCR8. (A) Cartoon depicting the functional domains of DGCR8 (NLS, nuclear localization signal; Rhed, RNA-binding heme domain; WW, WW domain; dsRBD1 and dsRBD2, double-stranded RNA binding domain 1 and 2). (B) HEK293T cells were transfected with plasmids overexpressing FLAG-tagged wild-type DGCR8 (DGCR8 FL), or DGCR8 lacking Drosha-interacting region (DGCR8 1-726), or an empty vector as a control, and subjected to anti-FLAG immunoprecipitations followed by analyses of co-immunoprecipitated endogenous RRP6 or Drosha with specific antibodies. Anti-FLAG antibody used to detect FLAG-tagged protein (bottom panel). (C) HeLa cell transfected with same plasmids as (B), were subjected to immunofluorescence experiments In order to identify subcellular localisation of FLAG-tagged protein. Nucleolin (red) was used as a nucleolar marker, while anti-FLAG (green) was used to detect FLAG-tagged proteins, and DAPI (blue) was used to stain nucleus.

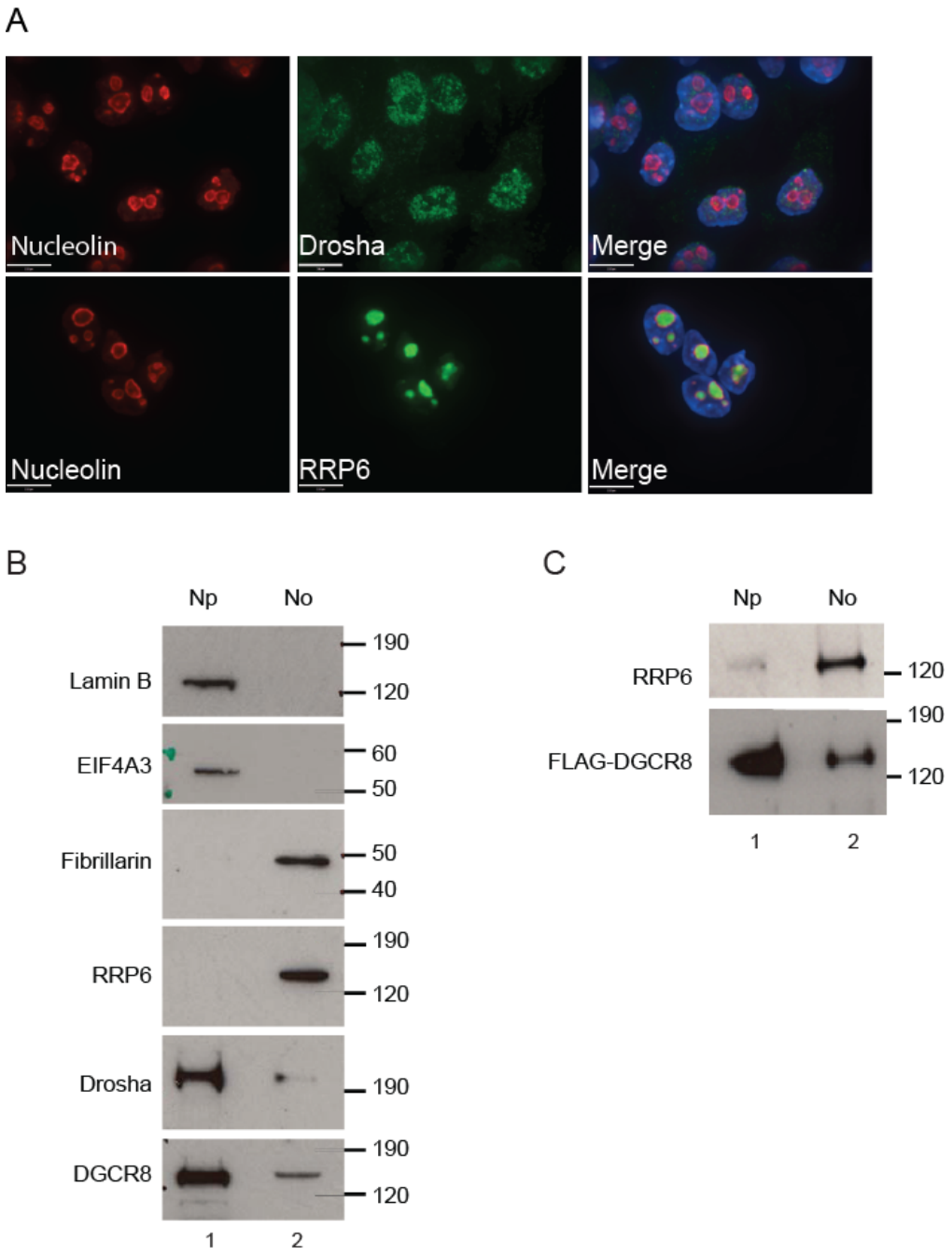


Figure 19 | DGCR8 is present within two nuclear compartments. (A) HeLa cells were subjected to immunofluorescence of endogenous RRP6 and Drosha (green). Endogenous Nucleolin serves as a nucleolar marker (red) whereas DAPI marks nuclear compartment (blue). (B) Nuclear fractionation of HeLa cells. Nucleoplasmic (Np, lane 1) and Nucleolar (No, lane 2) compartments are. Lamin B and EIF4A3 serve as nucleoplasmic markers. Fibrillarin serves as nucleolar marker. (C) Western blot analysis of anti-FLAG immunoprecipitations from nuclear fractionation of HeLa cells transiently transfected with FLAG-DGCR8. (B) Western blot analysis of the subcellular distribution of Drosha, DGCR8 and RRP6 in nucleoplasmic (Np, lane 1) and nucleolar fractions (No, lane 2). Lamin B and eIF4AIII served as nucleoplasmic markers whereas Fibrillarin is a nucleolar marker. Equivalent fractions of No and Np were loaded. (C) Western blot analysis of co-immunoprecipitated RRP6 with FLAG-DGCR8 from nucleoplasmic (lane 1) and nucleolar fractions (lane 2).

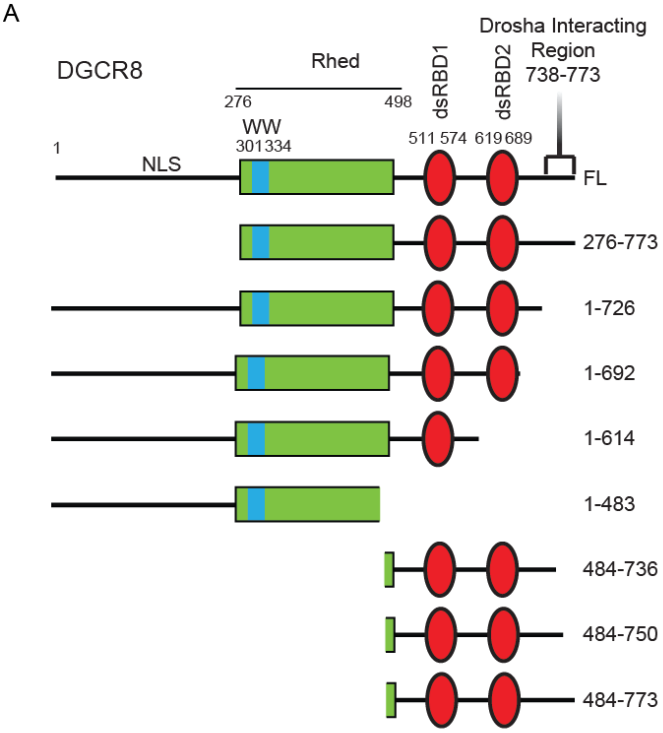
antibodies for immunofluorescence of the endogenous proteins. Interestingly, we found that the majority of overexpressed DGCR8 localised within the nucleoli but also within the nucleoplasm, independently of the tag used (Fig. 18 C, see FxDGCR8 FL; Fig. 22 C, see T7-DGCR8 FL). In order to confirm this result at the level of endogenous DGCR8 protein, we performed western blot analyses of nuclear fractionated HeLa cells. As fractionation controls we employed eIF4A3, a subunit of the spliceosome, and Lamin B, a protein integrated into the nuclear membrane as nucleoplasmic markers; and Fibrillarin, a snoRNP component, as a nucleolar marker. We observed faithful enrichment of subcellular markers within their appropriate fractions by western blot, confirming a good fractionation preparation (Fig. 19 B). In addition, we witnessed enrichment of RRP6 within the nucleolar fraction, whereas Drosha was predominantly present within the nucleoplasmic fraction (Fig. 19 B), in agreement with the immunofluorescence experiments (Fig. 19 B). Interestingly, we found DGCR8 present both within the nucleoplasm and nucleolar fractions, but being more enriched in the nucleoplasmic fractions (Fig. 19 B). To identify if subcellular localisation is a critical factor for the formation of the different DGCR8-ribonuclease complexes, we repeated nuclear fractionation on HeLa cells transfected with FLAG-DGCR8, followed by FLAG immunoprecipitation and RRP6 western blot analyses. We found that DGCR8 pulled-down RRP6 when localised to the nucleolus (Fig. 19 C). Altogether, this strongly suggests that DGCR8 localisation is an important factor for differential complex formation.

4.2.4 The dsRNA binding activity of DGCR8 is required for nucleolar localisation & RRP6 association

To define the region/s of DGCR8 required for RRP6 interaction we performed immunoprecipitation of endogenous RRP6 from cells transiently transfected with V5-DGCR8 truncated mutants, as depicted on Figure 20 A. Only full-length DGCR8 (V5-DGCR8 FL) and DGCR8 1-692 were efficiently co-immunoprecipitated with endogenous RRP6 and mutant DGCR8 276-773 was bound to a lesser extent (Fig. 21 A; lanes 2, 5 and 6, respectively). In addition, we confirmed that Drosha co-purified with V5-DGCR8 FL,

276-773, 484-750 and 484-773, validating previous findings (Fig. 21 B; top panel, lanes 2, 6, 8 and 9, respectively)(Yeom et al., 2006). As a control for these experiments, we performed immunoprecipitation using IgG, which did not co-purify any V5-tagged proteins, as expected (Fig. 21 B middle panel IgG). A summary of these results, together

A



B

DGCR8	RRP6 interaction	Drosha interaction	Subcellular * localization
DGCR8 FL	+++	+++	Np/No
D8 1-483	-	-	Np
D8 1-614	-	-	Np
D8 1-692	+++	-	Np/No
D8 276-773	+	+++	Np/No
D8 484-736	-	-	N/D
D8 484-750	-	++	N/D
D8 484-773	-	+++	Np/Cyt

Figure 20 | Immunoprecipitation experiments with DGCR8 truncated mutants identify region required for RRP6 interaction. (A) Cartoon depicting the functional domains of DGCR8 (NLS, nuclear localization signal; Rhed, RNA-binding heme domain; WW, WW domain; dsRBD1 and dsRBD2, double-stranded RNA binding domain 1 and 2). Schematic of truncated mutants are depicted below. All DGCR8 mutant are N-terminally tagged with a V5 epitope and the truncated mutant N- and C- terminal amino acids are displayed to the right. (B) Table summarising RRP6 and Drosha interaction with various DGCR8 mutants from westerns in Fig. 21 & 22. (*) a summary of cellular localisation (Cyt, cytoplasmic; Np, Nucleoplasmic; No, Nucleolar; N/D, not defined) of various DGCR8 mutants which was previously characterised by Yeom et al (2006).

with the previously published localisation studies are seen in Figure 20 B (Yeom et al., 2006). It is interesting to note that RRP6 can only interact with V5-DGCR8 mutants that have been previously shown to have nucleolar localisation; which, intriguingly, is only retained in those DGCR8 deletion constructs that harbour the two dsRBDs motifs (1-483, 1-614 & 1-692) (Fig. 20 A; Fig. 20 B red). In order to further validate this observation, we generated a mutant version of DGCR8, containing two point mutations within each dsRBD (T7-DGCR8-dsRBD; AA-KK & AS-KK) that abolished their RNA binding activity (Fig. 22 A). We found that DGCR8 RNA binding ability is required for its interaction with RRP6 and core components of the exosome, whereas Drosha binding is

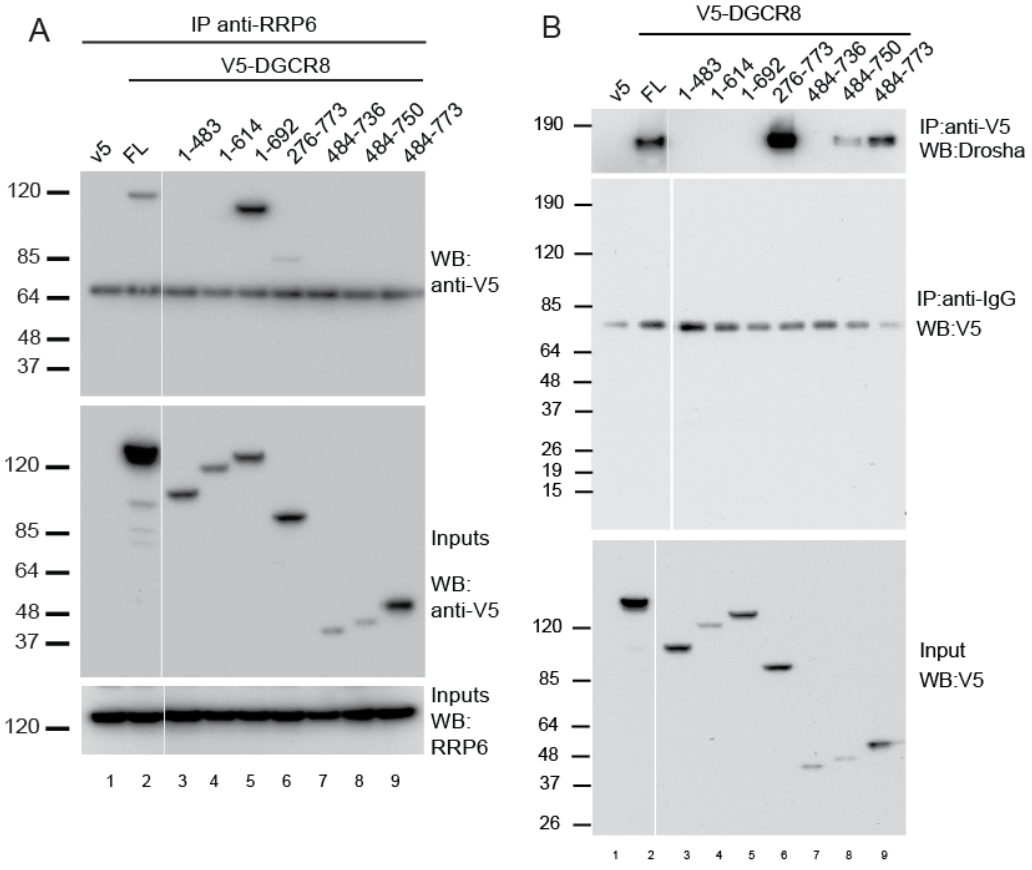


Figure 21 | Westerns from Fig20. HEK293T cells were transiently transfected with V5-tagged wild-type DGCR8 (FL), empty plasmid as a negative control (V5) and the indicated DGCR8 truncations depicted in Fig. 20 A. (A) immunoprecipitation of using anti-V5 antibody (top panel) or IgG antibodies as a control (middle panel), followed by Western blot with anti-Drosha antibody (top and panel). Inputs for V5-tagged proteins are shown in the bottom panel. (B) Cells expressing the same constructs detailed above were subjected to immunoprecipitation of endogenous RRP6 (bottom panel), followed by Western blot with anti-V5 antibody (top panel). Inputs are shown in the middle panel.

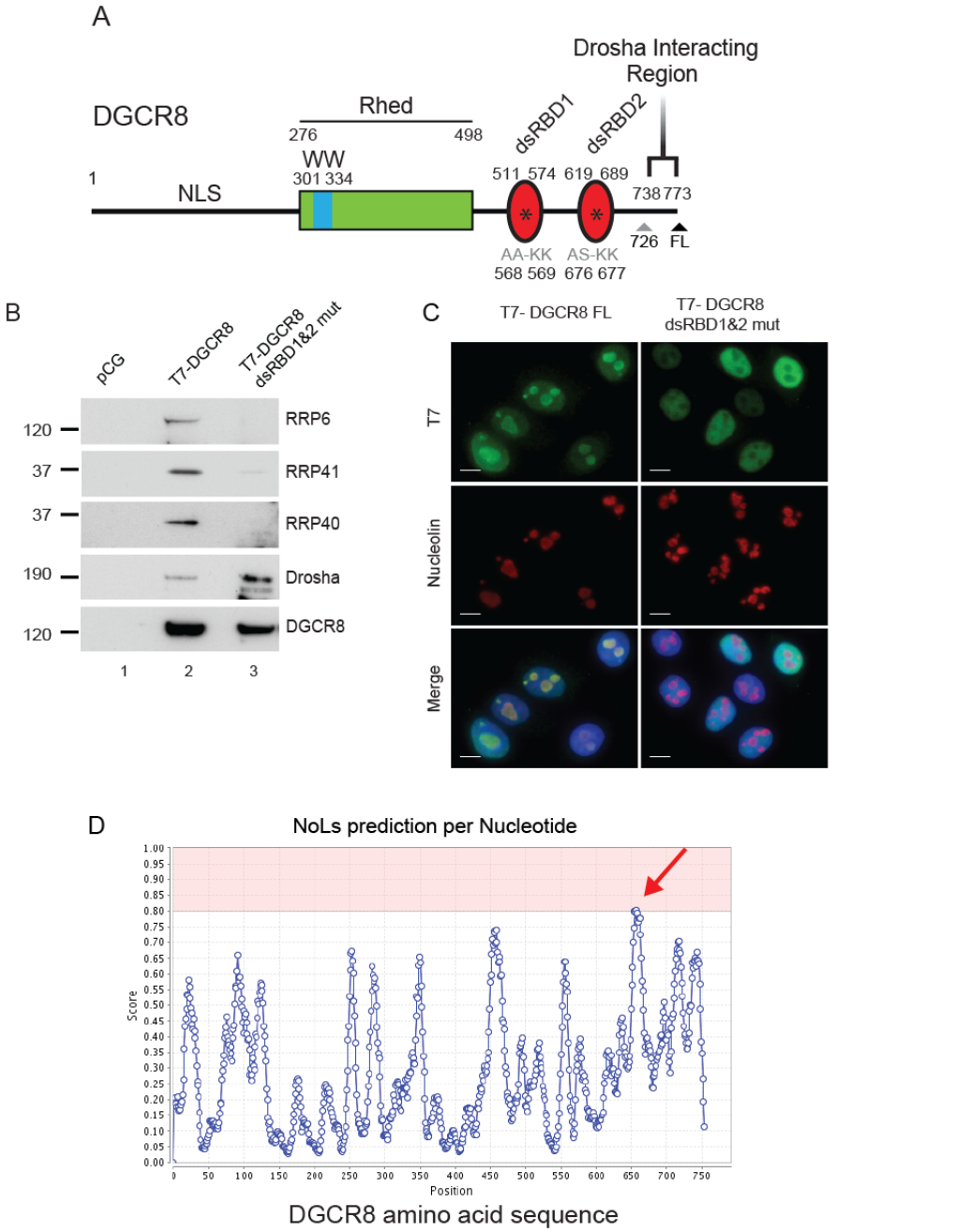


Figure 22 | The dsRNA binding activity of DGCR8 is important for nucleolar localisation. (A) Cartoon depicting mutated dsRBD (*) and below indicates the tandem amino acid mutations and their positions (NLS, nuclear localization signal; Rhed, RNA-binding heme domain; WW, WW domain; dsRBD1 and dsRBD2, double-stranded RNA binding domain 1 and 2). (B) Western blot analysis of T7-DGCR8 fl and T7-DGCR8 dsRBD1&2mut precipitates. pCG served as negative control (C) Indirect immunofluorescence analysis of T7-tagged versions of DGCR8 (green). Antibodies used indicated above. Anti-Nucleolin served as nucleolar marker (red) whereas DAPI stained nuclear compartment (blue). (D) Nucleolar Localisation signal motif prediction scores for DGCR8 protein sequences. Significance scores cross red region. Protein sequence of DGCR8 with annotated domains can be found in the appendix of this thesis. Nucleolar localisation signal found within dsRBD2.

unaffected (Fig. 22 B). We also observed that the subcellular localisation of T7-DGCR8-FL and T7-DGCR8-dsRBD was different. Whereas full-length DGCR8 could be found within the nucleoplasm and nucleolus, T7-DGCR8-dsRBD was exclusively nucleoplasmic (Fig. 22 C). Finally, we searched and identified a putative Nucleolar Localisation signal (NoLs) in DGCR8 primary protein sequence using an algorithm from the Barton lab (Scott et al., 2011). Importantly, this region lies within amino acids 657-676, which overlaps dsRBD2 in DGCR8 and specifically contains one of the amino acids mutated in T7-DGCR8-dsRBD (Fig. 22 D; for annotated sequence of DGCR8 see appendix), however point mutational analysis of this putative NoLs is require to disseminate if the sequence or rather the structure/function of the dsRBD permits nucleolar localisation. All of the above indicate that DGCR8 dsRNA binding activity is required for nucleolar localisation and RRP6 interaction.

4.2.5 DGCR8 acts as an dsRNA adaptor for RRP6

The identification of the exosome as a DGCR8 interacting partner in the nucleolus led us to ask whether DGCR8 could be acting as an adaptor for the recruitment of the exosome to snoRNAs. First, we tested binding of endogenous DGCR8 and RRP6 to mature and precursor snoRNAs (host pre-mRNA) by immunoprecipitation followed by qRT-PCR analysis (IP-qRT-PCR), using primers as depicted in Figure 23 A. We observed that DGCR8 and RRP6 did indeed associate to a similar extent with two representative C/D and H/ACA mature snoRNA molecules (mU16 and mU92, respectively) (Fig. 23 B). Interestingly, a subtle enrichment of host U16 and U92 pre-mRNA was demonstrated with DGCR8, which was not seen with RRP6 (Fig. 23 B), and as expected, we also detected binding of DGCR8, but not of RRP6, to a canonical Microprocessor substrate (pri-miR-24), whereas binding to 7SK was used as a negative control (Fig. 23 B).

We then further evaluated if the association between DGCR8 and snoRNAs was dependent on the ability of DGCR8 to bind RNA. For this purpose, we performed IP-qRT-PCR on overexpressed epitope tagged versions of DGCR8; T7-DGCR8-dsRBD and a wild-type version of DGCR8 (T7-DGCR8-FL). We confirmed that wild-type

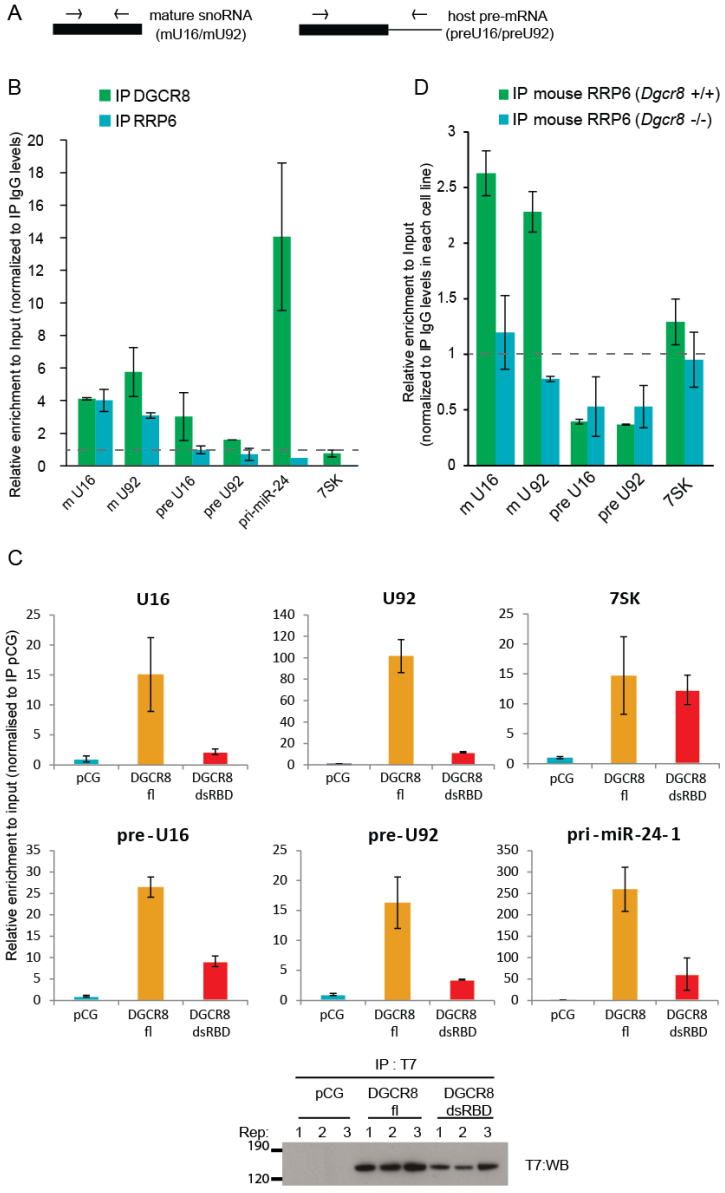


Figure 23 | DGCR8 acts a dsRNA adaptor; mediating snoRNA – RRP6 interactions. (A) Schematic representation of primer pairs used to amplify mature snoRNAs (black box) or host pre-mRNAs (line represents the host intron) by qRT-PCR. This approach is used by all experiments detailed in this figure. (B) Analysis of the associated mature snoRNAs (mU16 and mU92) and host pre-mRNAs (preU16 and preU92) to immunoprecipitated endogenous DGCR8 and RRP6 in HEK293T cells. Values represented show at least two different biological replicates with standard deviation. The enrichment of each RNA species is expressed relative to the levels of amplification in the control IgG immunoprecipitation (set to 1, represented as horizontal dashed line) and normalized to the levels in the Input material. (C) Analysis of the associated mature snoRNAs (U16 and U92) and pre-snoRNAs (pre-U16 and pre-U92) to overexpressed wild-type T7-DGCR8 and T7-DGCR8 dRBD1&2 mutant by qRT-PCR. Again, Pri-miR-24 serves as a positive control for a DGCR8 bound RNA and 7SK as a negative control. Inputs of immunoprecipitations are shown at the bottom. (D) Analysis of mouse RRP6 association to snoRNAs in mouse embryonic stem cells in the presence (Dgcr8 +/+) or absence (Dgcr8 -/-) of DGCR8 by qRT-PCR. (B,C,D) Quantitative RT-PCR data represents at least two different biological replicates with standard deviation.

overexpressed DGCR8 interacts with both mature and precursor snoRNAs tested, and that this interaction was direct, as enrichment of these RNA species was lost or severely reduced in the dsRBD mutant precipitates (Fig. 23 C). Again, we used 7SK as a negative control and confirmed that no significant enrichment was detected between FL and dsRBD DGCR8 (Fig. 23 C).

Taken together, these results suggest that DGCR8 could be the factor that enables RRP6 binding to snoRNAs, acting as an adaptor protein to efficiently recruit the exosome complex to these species. In order to test this possibility, we immunoprecipitated endogenous RRP6 protein from wild-type mouse embryonic stem cells (mESC) (*dgcr8* +/+) or the same cell line lacking DGCR8 (*dgcr8-/-*) and analysed associated RNAs by IP-qRT-PCR. Crucially, we found that mouse RRP6 binding to snoRNAs was abrogated in the absence of Dgcr8 (Fig. 23 D). This shows that DGCR8 is essential to promote binding of RRP6 to snoRNAs and suggests that DGCR8 acts as an adaptor to recruit the exosome complex to mature snoRNAs

4.2.6 DGCR8 and RRP6 regulate mature forms of snoRNAs

We next wanted to study if both DGCR8 and RRP6 are involved in the regulation of mature snoRNAs. To this end, we performed RNAi knockdowns experiments, in which relative levels of RNA were measured by qRT-PCR using primers specific to the mature snoRNA or host-mRNA sequence (Fig. 24 A). Relative levels of mRNA after 72-96 hour depletion are shown in Figure 24 D for HeLa cells and Figure 25 B for SH-SY5Y cells. Specifically, the abundance of mature U16 snoRNA increased upon single or tandem depletion of DGCR8 and RRP6 in HeLa cells to a similar extent; since they do not display additive effects this indicates that DGCR8 and RRP6 are working in a shared common pathway (Fig. 24 B). These observations were also recapitulated in SHY5Y neuronal cells and *dgcr8* -/- mESC (Fig. 25 A & C). Additionally, depletion of DIS3, the other associated exonuclease with the exosome core, did not display any significant changes (Fig. 24 B).

To dissect the roles of other adapter complexes within mature snoRNA regulation, we depleted RBM7 and ZCCHC7, which are subunits of the NEXT and hTRAMP complexes respectively; and hMTR4, a cofactor that facilitates RRP6 decay and a subunit of both the NEXT and hTRAMP complexes. Interestingly, depletion of these factors displayed opposing effects on mature snoRNA levels. Depletion of hMTR4, led to increase in the abundance of mature forms of U16; whereas depletion of ZCCHC7 or

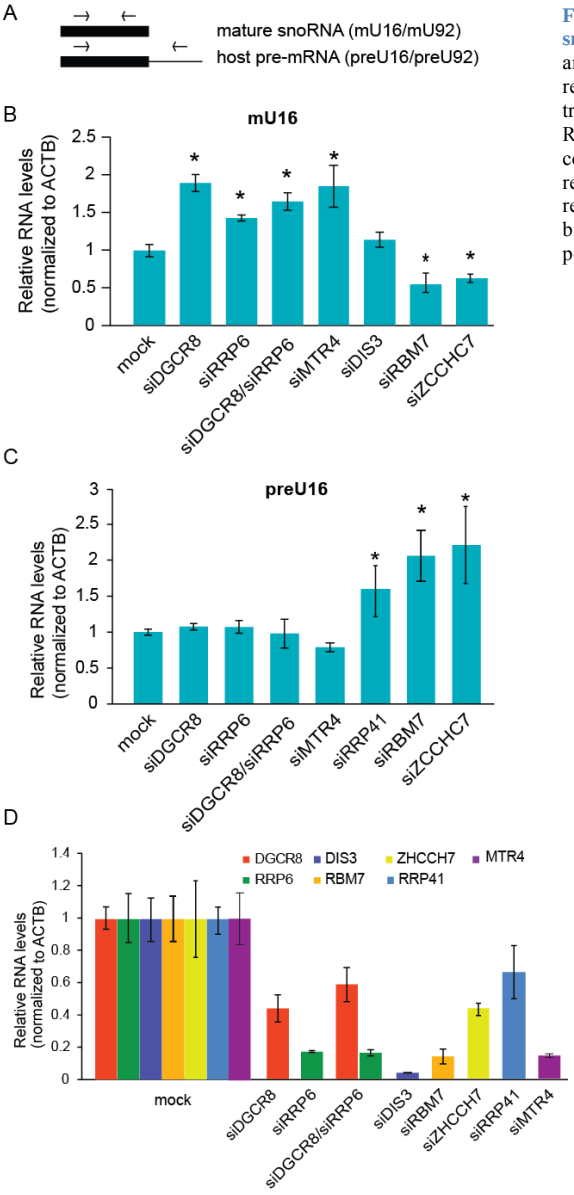


Figure 24 | DGCR8 and RRP6 control abundance of mature snoRNAs. (A) Schematic representation of primer pairs used to amplify mature snoRNAs (black box) or host pre-mRNAs (line represents the host intron) by qRT-PCR. (B,C,D) HeLa cells were transiently depleted of DGCR8, RRP6, MTR4, DIS3, ZCCHC7, RBM7 and RRP41. 72-96 hr after first transfection RNAs were collected and levels of mature U16 (B); pre-U16 (C); or their respective mRNAs (D), were assayed by qRT-PCR. All values represented in the two panels are the average of at least three biological replicas showing +/- s.e.m. Asterisks denote significant p-value (≤ 0.05) by Student's t test.

RBM7 lead to decreases in the abundance of mature U16 (Fig. 24 B). These data suggest, hMTR4 is important for regulating the abundance of mature snoRNAs whereas the NEXT and hTRAMP are required for snoRNA maturation. We next measured precursor host-mRNA levels of U16 snoRNA and observed that DGCR8, RRP6 and hMTR4 depletion did not show a significant accumulation of precursor forms of U16 snoRNA, whereas, ZCCHC7 and RRP41 knockdown caused an increased in these forms (Fig. 24 C).

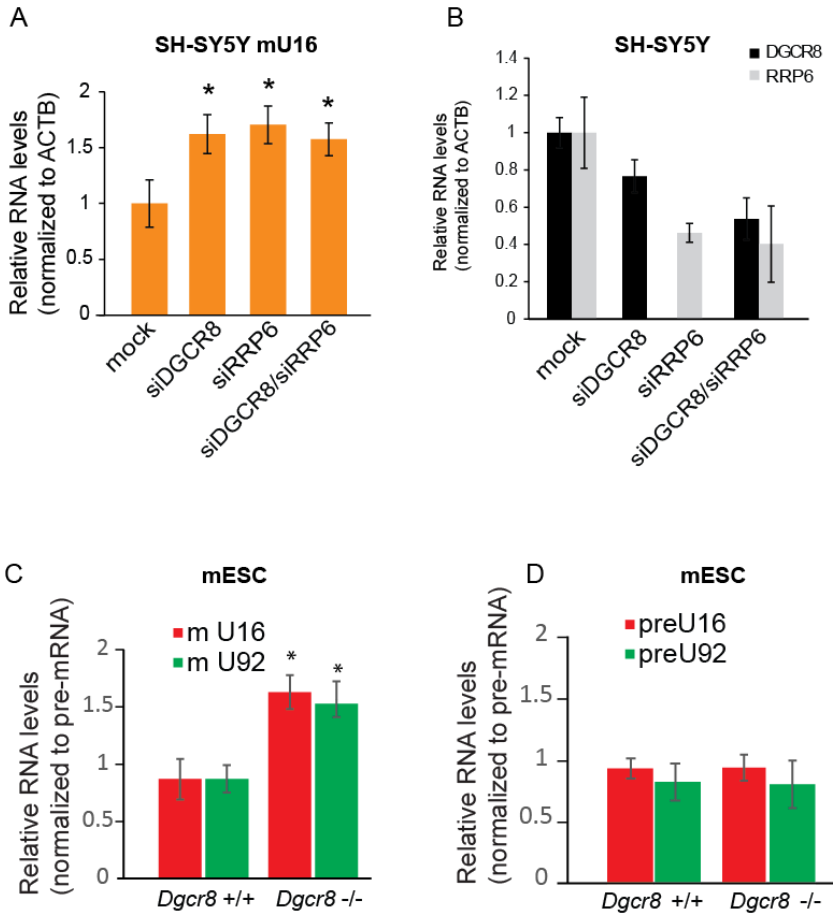


Figure 25 | Depletion of DGCR8/RRP6 in SH-SY5Y and mESC increases mature U16 and U92 levels. Same primer sets were used as depicted in Fig. 24 A to amplify mature snoRNAs (black box) or host pre-mRNAs (line represents the host intron) by qRT-PCR. (A,B) SH-SY5Y cells were transiently depleted of DGCR8 and RRP6. 72-96hr after first transfection RNAs were collected and levels of mature U16 (A); or their respective mRNA levels (B), were assay by qRT-PCR. (C,D) RNA was collected from Dgcr8 +/+ or Dgcr8-/- mESC and levels of mature U16 or mature U92 (C); or their precursor forms (D) were assayed by qRT-PCR. All values represented in the two panels are the average of at least three biological replicas showing +/- s.e.m. Asterisks denote significant p-value (≤ 0.05) by Student's t test.

4.3 DISCUSSION

4.3.1 Summary

In this chapter, we have further investigated the interaction between DGCR8 and RRP6-associated exosome complex identified in the previous chapter. We have identified DGCR8 acts as a dsRNA adaptor for the exosome by facilitating the interaction between snoRNAs and RRP6. We found the functions of DGCR8, as a dsRNA adapter protein, seem to be functionally distinct from the putative hTRAMP and NEXT complex as depletion of ZCCHC7 and RBM7 lead to decreased levels of mature snoRNAs. We also observed that depletion of either RRP6, DGCR8 or MTR4 led to accumulation of mature snoRNAs, whereas precursors were unaffected, which suggests that this complex is involved in the turnover of mature snoRNAs. We demonstrate that DGCR8 requires the dsRNA binding activity for nucleolar localisation and RRP6 interaction, which might be directed by a putative nucleolar localisation signal embedded within the second dsRNA binding domain of DGCR8. Finally we demonstrated the regulation by DGCR8/RRP6 complex on mature snoRNAs is independent of class or genomic context. The biological relevance is discussed further in chapter 7.

4.3.2 The RRP6 interacting region and nucleolar localisation of DGCR8

Testing of various DGCR8 deletion mutants with RRP6 highlighted the requirement for intact dsRBD activity of DGCR8 to efficiently interact with RRP6, which was also reported to be important for interactions with MeCP2 and Nucleolin (Cheng et al., 2014; Shiohama et al., 2007). We also found that ablation or inactivation of the dsRBDs resulted in exclusive nucleoplasmic DGCR8 localisation. Accordingly, compartmental specific IPs show DGCR8 can only co-purify RRP6 within the nucleolus. All of which advocates the formation of these complex is dependent on the localisation of both factors. In addition, the comparisons of DGCR8 deletion mutants specify a region between amino acids 267-692 that was consistently present in mutants able to interact with RRP6. This region contains the dsRBD and the Rhed motifs, which features a WW proline

rich domain that has been shown to be important for heme binding and dimerisation (Faller et al., 2007; Quick-Cleveland et al., 2014). However, to truly identify the RRP6 interacting region, *in vitro* binding assays will need to be performed using purified DGCR8 deletion mutants and RRP6, if these interaction are through direct contact between these factors. If, on the other hand, these interactions were shown not to be direct, a likely candidate to bridge this interaction is MTR4, which has been identified in every adaptor complex to date and also interacts with DGCR8 (data not shown). Functional studies within yeast and human cells have evidenced that deficiencies in Mtr4/MTR4 or the exosome complex caused increased abundances of a similar profile of precursor RNAs (de la Cruz et al., 1998; Lubas et al., 2011; Schilders et al., 2007). Furthermore, Rrp6p-Rrp47p mediate the interactions between Mtr4 and the exosome core (Schuch et al., 2014), which are also conserved in humans (Lubas et al., 2011; Schilders et al., 2007). Therefore, co-IP experiments of DGCR8 and RRP6 in cells transiently depleted of MTR4 will identify if RRP6 and DGCR8 interaction is bridged by MTR4.

In addition, it will be of particular interest to study the stoichiometry of DGCR8 protein within this novel complex. This could be done by reconstitution of the DGCR8-RRP6 complex *in vitro,* which will enable us to study the stoichiometry of these components and estimate their sizes using techniques such as native PAGE to analyse complex migration, and/or sedimentation analyses.

Chapter 5. Identification of the direct RNA targets of the RRP6/DGCR8 complex

5.1 INTRODUCTION

In the previous chapter we demonstrated a functional link between DGCR8 and the nuclear exosome. We found that DGCR8 acts as an adaptor protein, specifically targeting the RRP6-associated exosome complex to mature snoRNAs for their degradation. For the snoRNAs tested, this regulation was independent of their class (U16, C/D box; U92, H/ACA box), however, whether the RRP6/DGCR8 complex can function to degrade all snoRNAs and/or other cellular remains an open question.

To answer this, we reasoned that overlapping the RNA binding profiles of DGCR8 and RRP6 would identify common substrates, and hence potential direct RNA targets of the DGCR8/RRP6 complex. To this end, we identified the direct RNA targets of RRP6 using individual resolution Cross Linking ImmunoPrecipitation (iCLIP). As previously described, iCLIP is a method developed for the identification of direct RNA substrates of an RNA binding protein, that uses *in vivo* UV irradiation to form a covalent bond between RNA and proteins. This enables highly stringent and robust isolation of protein:RNA complexes (Konig et al., 2011).

The 3'→5' exoribonuclease RRP6 is required for processing and degradation of many cellular RNAs; as such, overexpression of dominant negative forms stabilise, usually lowly abundant, processing/degradation intermediates (Sloan et al., 2013). Structural and conservational analyses of the active site of RRP6 identified a highly conserved DEDD-Y motif, in which, the substitution of the first aspartic acid (D) to asparagine (N) abrogates exoribonuclease activity (Januszyk et al., 2011). To carry out the iCLIP analysis of RRP6, we transiently overexpressed both a FLAG-tagged version of wild-type (WT) and a catalytic mutant (CAT; D313N) in order to stabilise transient interactions with processed RNAs. We also performed iCLIP on endogenous (ENDO) RRP6 to complement these experiments.

Within this chapter, we have collaborated with Dr. Mireya Plass (University of Copenhagen) who mapped the hRRP6 iCLIP sequencing data, generated by myself, and also reanalysed DGCR8 CLIP (Macias et al., 2012). Furthermore, Dr. Mireya Plass

also identified significant RRP6 binding sites within the transcriptome and generated average iCLIP read density profiles shown within this chapter (5) and chapter 6. In addition, Dr. Sara Macias (University of Edinburgh) performed experiments displayed in Figure 31 E & F.

5.2 RESULTS

5.2.1 iCLIP of epitope tagged RRP6 WT and CAT

In total, two biological iCLIP libraries replicates for both RRP6 WT and CAT were produced. Firstly, HEK293T cells transiently expressing FLAG, FLAG-RRP6 WT or FLAG-hRR6 CAT were expanded in 15cm plates until 95% confluency. Next, each plate was treated with UV light in order to covalently crosslink protein:RNA complexes that interacting directly. Thereafter cells were collected, lysed and cellular RNAs were partially degraded using high and low concentrations of RNase (high, +++; low, +). Finally, we performed anti-FLAG immunoprecipitations and radiolabelled the precipitates, in order to visualise protein:RNA complexes (for brief schematic see Fig. 26 A). A more detailed protocol is included in the Materials and Methods section of this thesis.

Increased signal from radiolabelled RNA in UV-treated lanes vs non-UV-treated lanes identified RRP6 protein crosslinks to RNA (Fig. 26 B & C). Furthermore, IPs from lower RNase concentrations resulted in weaker signal that was distributed up from the size of the FLAG-RRP6 (120kDa), whereas higher RNase concentrations displayed a more focused signal around 120kDa (Fig. 26 B &C). It has been reported that proteins are also labelled by the polynucleotide kinases (PNKs) used in the iCLIP protocol (Huppertz et al., 2013; Konig et al., 2011). Indeed, we observe this in our autoradiograms in lanes with immunoprecipitations from samples expressing FLAG-RRP6 proteins (Fig. 26 B &C), importantly; we do not witness this in the negative FLAG IPs (Fig. 26 B &C; lanes 1 & 2). Altogether these data indicate a robust purification of FLAG-tagged proteins. Next, protein:RNA complexes were isolated above RRP6 migration (120kDa) from low RNase treated anti-FLAG IPs (Fig. 26 B & C; see red line) and were further treated with proteases. Following this, RNAs were isolated and reverse transcribed into cDNAs that were separated by PAGE and isolated at selected sizes (high, 120-200nt; medium, 120-85nt; low, 85-70nt) and further amplified. Finally, samples of the cDNAs amplified from high, middle and low size ranges were analysed by PAGE to ensure the libraries were amplified to their correct corresponding sizes after addition of Illumina adapters. Also the

low density of primer dimers (141nt) compared to cDNA signal indicates highly complex libraries were produced (Fig. 27 A & B). Thereafter, barcoded CLIP libraries were mixed and sequenced.

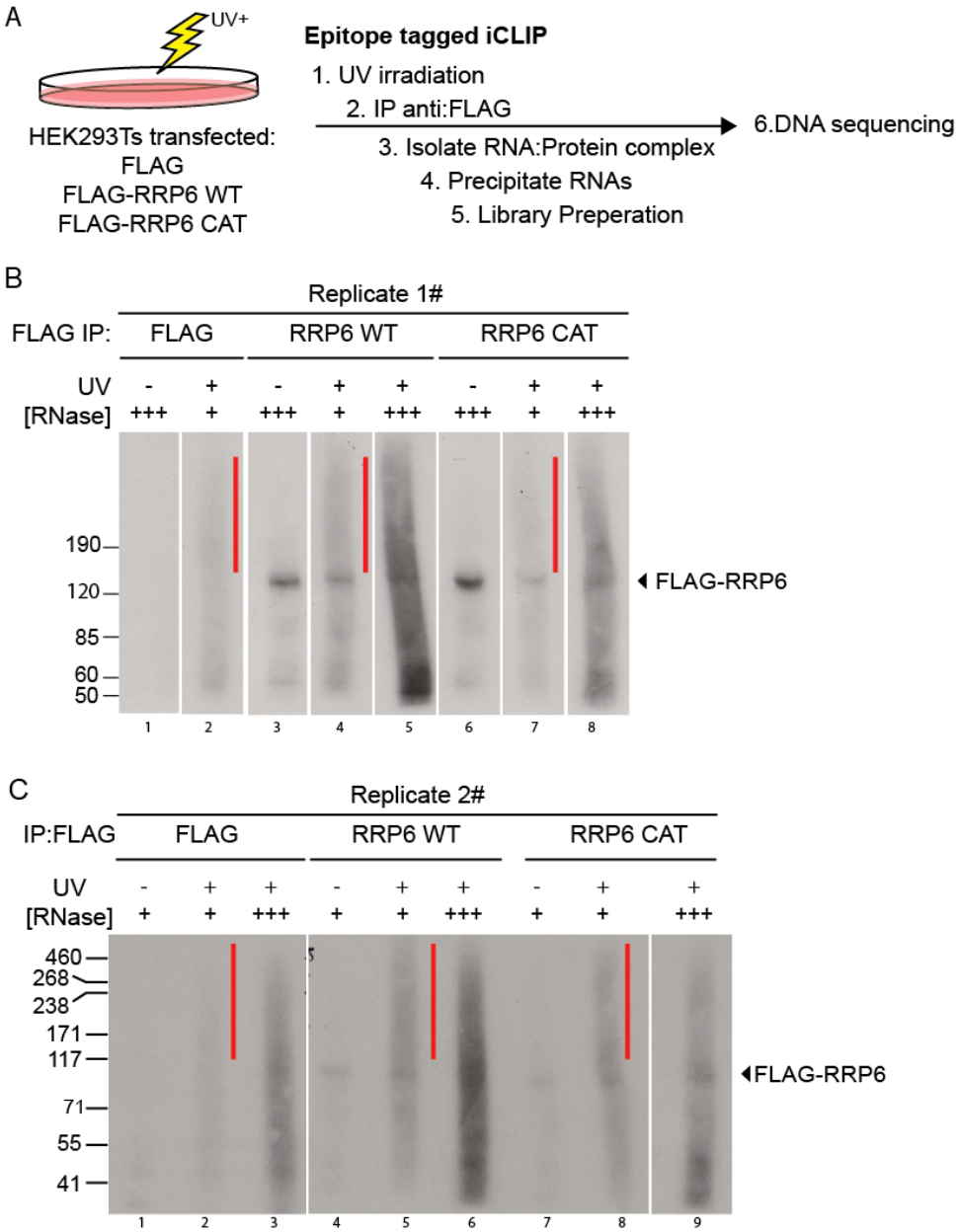


Figure 26 | Overexpressed FLAG tagged RRP6 WT and CAT iCLIP autoradiograms and cDNA gels of Illumina libraries. (A)Schematic of protocol followed for epitope tagged RRP6 iCLIPs. (B,C) Protein:RNA complex were radiolabelled and separated by SDS-PAGE. Protein:RNA complexes were isolated from low (+) RNase treated FLAG, FLAG-RRP6 WT, and FLAG-RRP6 CAT immunoprecipitations ([B; lanes 2,4,7][C; lanes 2, 5, 8]), in regions above the normal migration (red line) of uncrosslinked RRP6 (120kDa). Triangle denotes non-specifically radiolabelled protein. (+++) marks high RNase treated samples. Replicate 1 is shown in (B), whereas replicate 2 is displayed in (C)

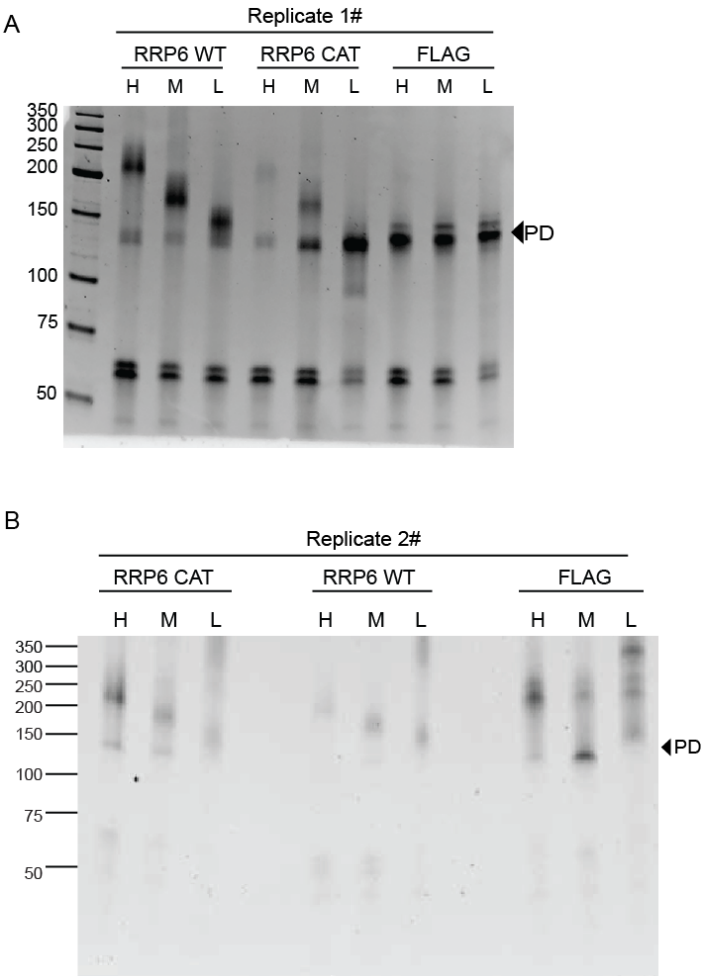


Figure 27 | cDNA epitope tagged. (A,B) Samples of Illumina libraries sent for sequencing. Replicate 1 from Fig. 26 A is displayed in (A), whereas replicate 2 from Fig. 26 B is displayed in (B). H, M and L denote high, middle and low length cDNAs isolated after reverse transcription of isolated RNAs. Primers run below 75nt, whereas primers dimers (PD) run at 141nt.

5.2.2 iCLIP of endogenous RRP6

In parallel, we also identified the RNA binding profile of endogenous RRP6 (ENDO). This approach was substantially challenging due to many factors. First, high salt conditions were required for stringent purification of protein:RNA complexes. Moreover, the relatively low levels of endogenous RRP6 combined with an average quality of RRP6 antibodies led to large amounts of antibody being required (10μg) and resulted in only one iCLIP library being produced.

To generate endogenous RRP6 iCLIP libraries we performed two independent IPs. The RNAs from these IPs were pooled following protein:RNA complex isolation and as a negative control we used species-matched IgG (for a brief summary see Fig. 28 A).

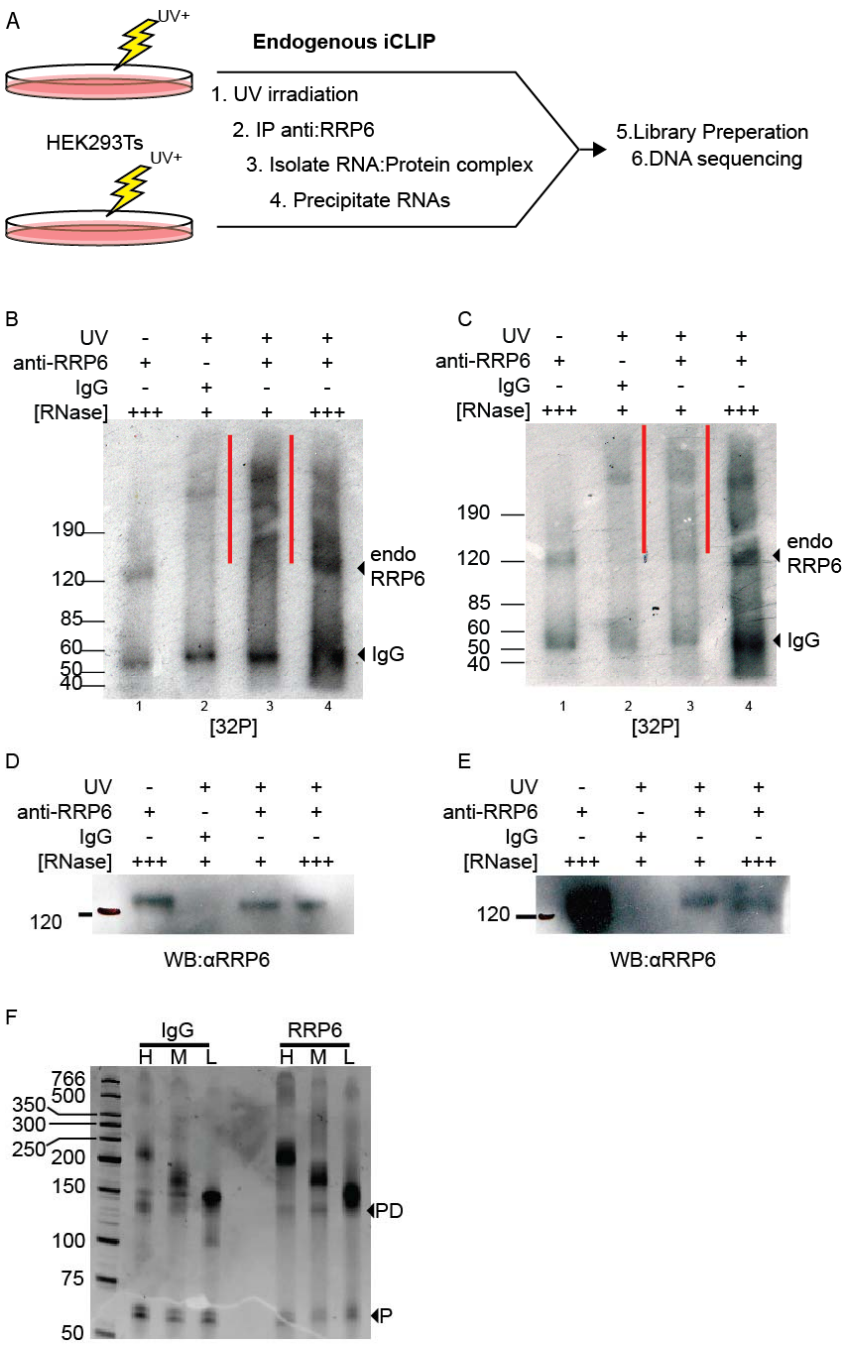


Figure 28 | Endogenous (ENDO) RRP6 iCLIP autoradiograms and cDNA gels of Illumina libraries. (A)Schematic of protocol followed for endogenous iCLIP of RRP6. (B,C) Protein:RNA complex were radiolabelled and separated by SDS-PAGE. Protein:RNA complexes were isolated from low (+) RNase treated anti-IgG ([B; lane 2][C; lane 2]) and anti-RRP6 ([B; lane 3][C; lane 3]) immunoprecipitations from regions above the normal migration (red line) of uncrosslinked RRP6 (120kDa). Antibodies used for IP indicated above autoradiogram (+/-). endo-RRP6 denotes non-specifically radiolabelled protein. (+++) marks high RNase treated samples.(D,E) Western blot analyses of 10% of samples used for IPs in (B) and (C), respectively. (F) Samples of Illumina libraries sent for sequencing. H, M and L, denote high, middle and low length cDNAs isolated after reverse transcription of isolated RNAs. Primers run below 75nt, whereas primers dimers (PD) run at 141nt.

Again, the specificity of the IP was shown by the signal around the migration size of RRP6 in lysates treated with high RNase concentrations (Fig. 28 B & C); and also no signal was detected in 10%-IP inputs IgG control for RRP6 (Fig. 28 D & E). IPs from lower RNase conditions displayed a more diffuse signal extending upwards of endogenous RRP6 migration (Fig. 28 B & C; lanes 2 and 3). Thus we isolate protein:RNA complexes from these regions, in addition to the same region in the IgG control (see red line in Fig. 28 B & C). The analysis of endogenous RRP6 libraries by PAGE showed correct sizes of amplified cDNAs after addition of Illumina adapters and low levels of primer dimers, however there was DNA amplification in the negative control, which means that there was some nucleic acid contamination during the purification (Fig. 28 F). We therefore decided to sequence libraries produced by both IgG and endogenous RRP6, in order to discriminate real substrates from common contaminants. Finally, barcoded iCLIP libraries were mixed and sequenced.

5.2.3 RRP6-ENDO, -WT and -CAT iCLIP libraries display strong correlations

The total number of reads obtained after sequencing, demultiplexing of barcodes and collapsing of randomised adaptors are shown in Table 7. Next, reads were mapped to hg19 using bwa-pssm (Kerpedjiev et al., 2014) and only unique reads with high mapping scores were analysed further. Biological replicates from both WT and CAT displayed highly reproducible counts within annotated genes (Pearson's correlation 0.55 and 0.44 respectively) (Fig. 29 A & B). Following this observation, we pooled biological replicates for subsequent analyses due to high reproducibility and in order to increase coverage of lowly expressed RNAs. Importantly, comparisons of pooled WT, CAT and ENDO read counts again identified high correlations (Pearson's correlations >0.78) (Fig. 29 C-E) indicating that WT, CAT and ENDO RRP6 share a very similar RNA binding landscape. Total numbers of significant clusters identified from pooled samples are displayed in Table 8.

5.2.4 Identification of significant RRP6 binding sites

We next identified RRP6 significant binding sites. As previously described, iCLIP analyses relies on the identification of truncated cDNAs that are produced by reverse transcriptase stalling at short peptides left at the UV-induced crosslink site (Fig. 14). Hence, protein:RNA interactions can be mapped at nucleotide resolution, by identifying enrichment of these cDNAs, followed by extrapolating one base upstream of

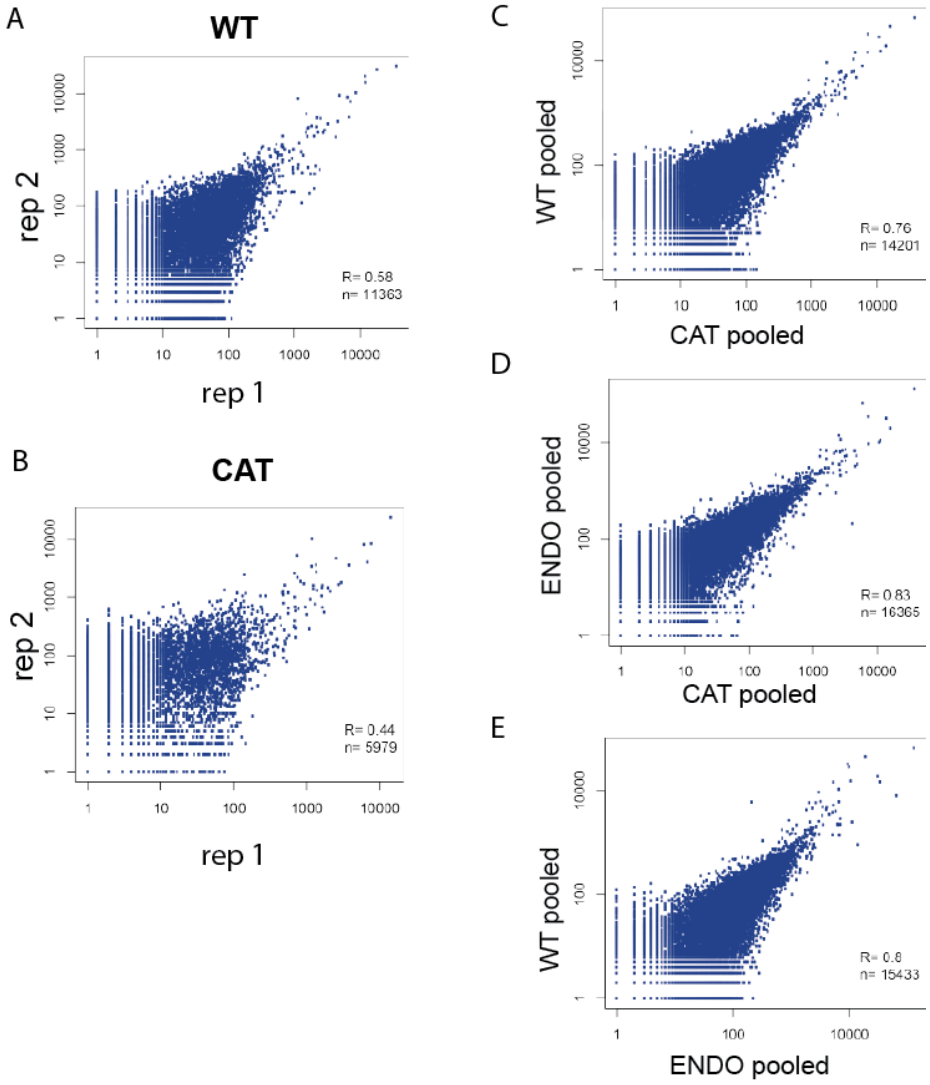


Figure 29 | RRP6 iCLIP experiment correlations. (A-E) After mapping reads to hg19, reads mapping annotate gene (Ensembl70) were calculated (R = Pearson's correlation; n = number of genes represented on graph). (A,B) Correlation of read counts for biological replicates of FLAG-RRP6 WT (A); or FLAG-RRP6 CAT(B). Following these analyses, biological replicates from FLAG-RRP6 WT and FLAG-RRP6 CAT were pooled for further analyses (C-E) Correlations of read counts for pooled FLAG-RRP6 WT and FLAG-RRP6 CAT (C); RRP6-ENDO and FLAG-RRP6 CAT (D); and, FLAG-RRP6 WT and RRP6-ENDO (E). These graphs are presented on a log10 scale.

5' of the sequenced read. These conventional iCLIP analyses are geared towards identifying consensus motifs for RBPs positioned within RNAs. In contrast, RRP6 is a processive 3'→5' exoribonuclease, which will be positioned at the 3' terminal end of its

Table 7 | Hierarchical break down of iCLIP sequencing results.

Sample	hRRP6 WT rep1		hRRP6 WT rep2	
Replicate	Counts	%	Counts	%
Final reads	32556116	100.00	34368264	100.00
Unique reads	2107322	6.47	2010991	5.85
Confidently mapped	764794	36.29	875916	43.56
genome	681833	32.36	795711	39.57
exome	78345	3.72	77367	3.85
Exon-junction	4616	0.22	2838	0.14

Sample	hRRP6 CAT rep1		hRRP6 CAT rep2	
Replicate	Counts	%	Counts	%
Final reads	15232740	100.00	30856132	100.00
Unique reads	849552	5.58	2855319	9.25
Confidently mapped	257083	30.26	924434	32.38
genome	221140	26.03	814145	28.51
exome	35130	4.14	106612	3.73
Exon-junction	813	0.10	3677	0.13

Sample	hRRP6 ENDO		IgG	
Replicate	Counts	%	Counts	%
Final reads	69012050	100.00	29083539	100.00
Unique reads	8023318	11.63	1281197	4.41
Confidently mapped	2735449	34.09	515975	40.27
genome	2387690	29.76	467227	36.47
exome	331256	4.13	45929	3.58
Exon-junction	16503	0.21	2819	0.22

Table 8 | Number of significant clusters identified from RRP6 iCLIP experiments.

	RRP6 WT pooled		RRP6 CAT pooled		RRP6 ENDO pooled		IgG	
	Counts	%	Counts	%	Counts	%	Counts	%
merged reads	164071		1181517		2735449		515975	
Clusters	123084	100	195378	100	890124	100	56581	100
Sig. Clusters	65801	53	86024	44	139997	15	28367	50

RNA targets. Thus, we reasoned that RRP6 will crosslink to the 3' end of the transcripts that is degrading, and depending on the extent of degradation, this could be anywhere throughout the annotated gene. Due to the nature of the iCLIP protocol, truncated cDNAs produced by this would be too small to sequence or map efficiently. Following this rationale, significant clusters were calculated using read densities and not 5' ends of cDNAs (for details see methods). In total we identified 65,801 significant clusters for WT; 86,024 for CAT, and 139,997 for ENDO, respectively (Table 8). The nature of these binding sites will be discussed in the next chapter (6).

5.2.5 Classification of DGCR8-dependent targets of RRP6

We defined a role for DGCR8 as a molecular adaptor required for the RRP6-mediated degradation of U16 and U92 snoRNAs. This led us to postulate that DGCR8/RRP6 could be responsible for the regulation of all snoRNAs and/or other cellular RNAs. In order to investigate this, we selected the overlapping significant clusters in RRP6 (ENDO and WT), and intersected these with DGCR8 CLIP clusters (for schematic see Fig. 30 A)(Macias et al., 2012), and the location of these overlapping clusters was then annotated. Using this approach we identified 358 non-coding RNAs (ncRNAs) as common substrates. Of these, the most represented class of ncRNAs was snoRNAs (166 snoRNAs) followed by tRNAs (91 tRNAs) (Fig. 30 B). To assess the coverage of binding of DGCR8 and RRP6 to snoRNAs, we first identified which snoRNAs are expressed within HEK293T cells, by employing a previously published small RNAseq dataset (Kishore et al., 2013). Of the 1,408 annotated snoRNAs, 422 were expressed in HEK293T cells. From the total number of HEK293T expressed snoRNAs, 47% of SNORD (120/255); 24% of SNORA (35/145) and 50% of SCARNA (11/22) were bound by both DGCR8 and RRP6 CLIP analyses (Fig. 30 B). Next, we calculated the average read density of CLIP tags from both RRP6 and DGCR8 over snoRNA genes on a genome wide context. We found that DGCR8 and RRP6 CLIP reads fall mainly within the annotated sequence of the snoRNAs and that this was common between all the snoRNA classes (Fig. 30 C), which means that they mainly bind to the mature sequence,

and not to the surrounding intronic sequences. In addition, CLIP read density profiles identified specific increased DGCR8 CLIP density within the 5' and 3' regions of SNORA and SNORD snoRNAs, whereas RRP6 was bound throughout all families of snoRNAs. This could reflect that DGCR8 is binding to specific regions of snoRNAs; whereas the

RRP6 binding sites identified by the CLIP analysis, illustrates that RRP6 is bound throughout the entire snoRNA sequence, as it is actively degrading it. Altogether, these data indicate DGCR8 and RRP6 can bind a wide range of mature snoRNAs, suggesting the DGCR8/RRP6 complex can regulate the abundance of mature snoRNAs on a global scale.

5.2.6 The DGCR8 /RRP6 complex regulates steady state levels of Human Telomerase RNA component (TERC/hTR)

Human Telomerase RNA (hTR), is the RNA component of the human telomerase enzyme and features a H/ACA box near its 3' end (Fig. 31 A) that is essential for pre-RNP formation and nucleolar localisation of the telomerase RNP itself (Jády et al., 2004; Mitchell et al., 1999; Pogacić et al., 2000). CLIP data also revealed that both DGCR8 and RRP6 bind to hTR, with DGCR8 binding being mostly concentrated towards the 3'end, where the H/ACA domain is located (Fig. 31 B). We confirmed that overexpressed DGCR8 associates with hTR by IP-qRT-PCR, and these interactions were specific, since this association was abolished when using a dsRBD mutant of DGCR8 (T7-DGCR8 dsRBD) (Fig. 31 C). Importantly, depletion of DGCR8, RRP6 or a combination of both resulted in hTR accumulation in HeLa cells, whereas no changes in hTR levels were observed upon Drosha depletion (Fig. 31 D). Interestingly, an increase in the levels of hTR/TERC has been reported to boost telomerase enzyme activity leading to inappropriate lengthening of telomeres (Cristofari and Lingner, 2006), therefore we decided to quantify telomere length in the absence of DGCR8. To note, these experiments discussed next were performed by Dr. Sara Macías. First, she confirmed that DGCR8 is also controlling TERC RNA levels in mESC (*Dgcr8* -/-), whereas mESC lacking Dicer exhibited no increase in TERC levels; confirming that changes in TERC levels are not an

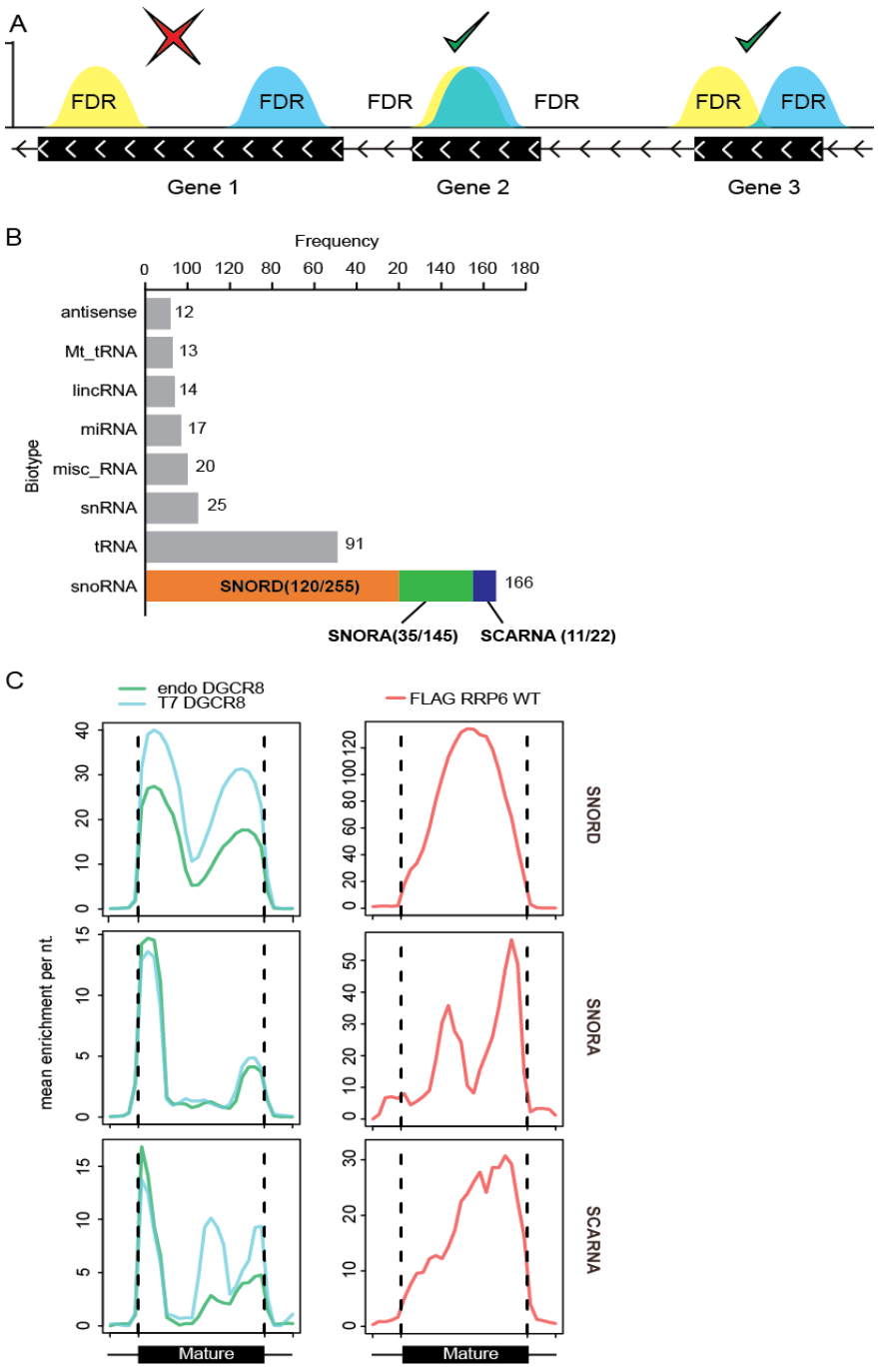


Figure 30 | Common RNA substrates of DGCR8 and RRP6 identified by overlapping CLIP profiles. (A) Schematic showing strategy to identify common substrates. (B) Common substrates of RRP6 and DGCR8 identified using strategy defined in (A). Numbers located inside snoRNA histogram identify the number of individual snoRNAs bound by both RRP6/DGCR8 compared to number expressed in HEK293T cells. "Expressed snoRNAs calculated from sRNAseq data set from Kishore et al. (2013). (C) Average read density of FLAG-RRP6-WT iCLIP and DGCR8 CLIP Macias et al. (2012) over SNORD,SNORA and SCARNA classes of snoRNAs. Mature sequence split into 20 bins. 5' and 3' flanks are 50nt up and downstream, respectively of the mature sequence, these regions are split into 5 bins (see methods for more details on calculation).

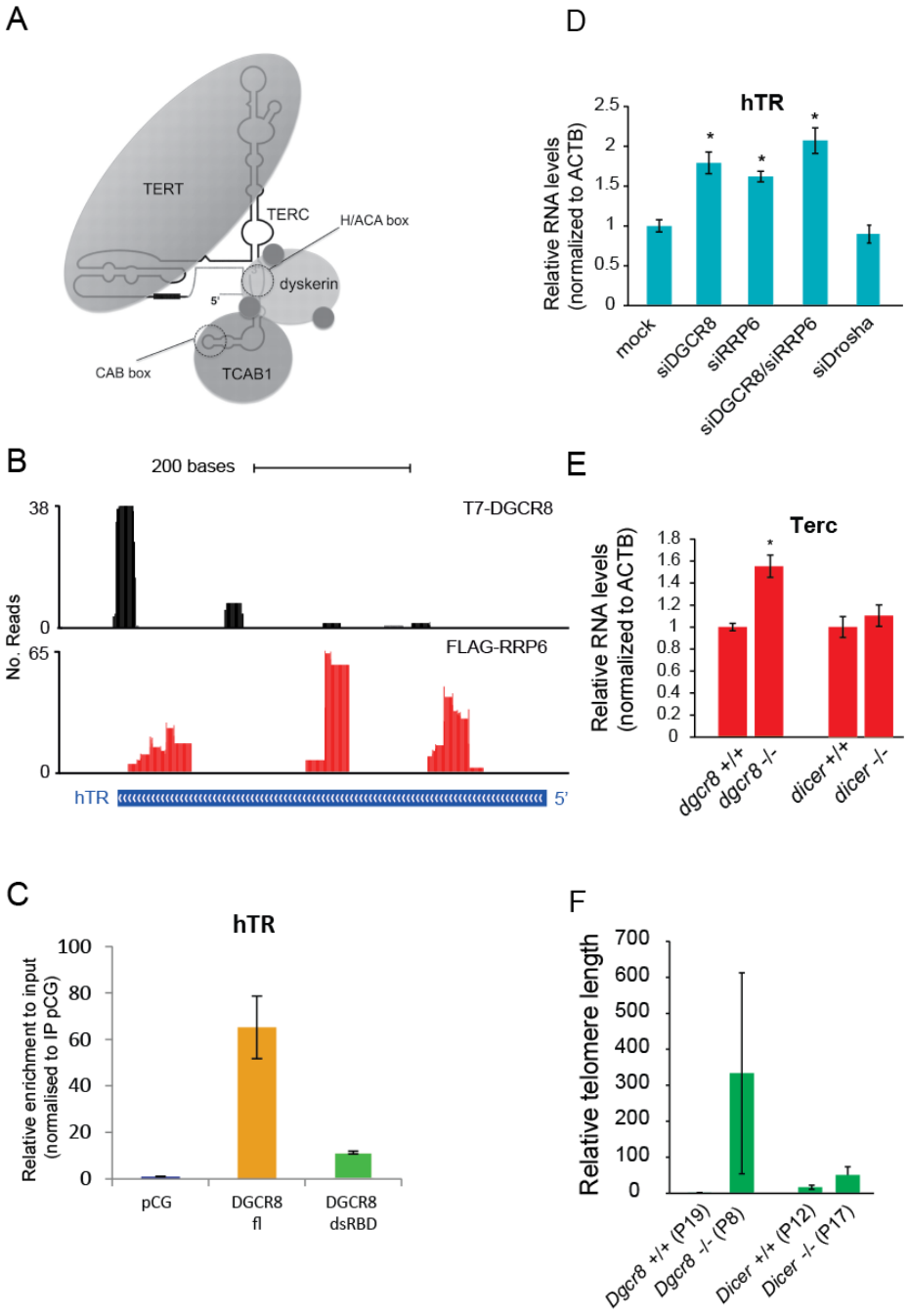


Figure 31 | The H/ACA box containing hTR/TERC is regulated by RRP6 and DGCR8. (A) Cartoon of hTR/TERC structure with H/ACA box in its 3' end annotated (taken from Artandi and DePinho [2009]). (B) Read density plots of T7-DGCR8 CLIP (Macias et al., 2012) and RRP6 WT iCLIP over the hTR loci. (C) IP-qRT-PCR analyses of hTR levels in HEK293T cells transiently overexpressing T7 tagged-DGCR8 fl (wildtype), a double stranded mutant (DGCR8 dsRBD; for schematic see Fig. 22 A) and pCG (negative control). For inputs please see Fig. 23 C. (D) qRT-PCR analyses of hTR levels in HeLa cells depleted of DGCR8, RRP6 or Drosha. (E,F) mESC devoid of Dgcr8 or dicer and their respective wild type parental cell lines were subjected to; qRT-PCR to measure TERC RNA levels (E); or qPCR to measure the number of telomeric repeats, which is an indirect assay to calculate telomere length (Callicott and Womack, 2006). Please note experiments (E) and (F) were performed by Dr. S. Macias.

indirect effect of perturbation of miRNA biogenesis (Fig. 31 E). To examine the consequences of increased abundance of hTR/TERC, she analysed telomere length in the absence of DGCR8 and Dicer via an indirect assay that measures telomeric repeats using qPCR (Callicott and Womack, 2006). She confirmed previous reports where the absence of Dicer resulted in a moderate increased in telomere length (Fig. 31 F). Strikingly, the absence of DGCR8 also boosted telomere length significantly more than absence of Dicer (Fig. 31 F); indicating that loss of miRNA and DGCR8 activity on controlling telomerase RNA levels may have an additive effect on telomeric repeats; and further highlighting the miRNA biogenesis independent functions of DGCR8. Altogether, these data suggests that the DGCR8-RRP6 alternative complex may have a role in controlling the number of telomeric repeats by regulating hTR levels.

5.3 DISCUSSION

5.3.1 Summary

The overarching aim from these series of experiments was to combine the RNA-binding profiles of DGCR8 and RRP6 to define common substrates for these proteins. This enabled us to ascertain that the DGCR8/RRP6 complex can bind RNA substrates other than U16 and U92 snoRNAs. Our main finding from these analyses is that snoRNAs were the most highly represented class of ncRNA, followed by tRNAs. The binding of DGCR8 and RRP6 to snoRNAs was independent of class and genomic context, raising the interesting possibility that DGCR8/RRP6 can regulate all bound snoRNAs. Furthermore, we could identify the human telomerase RNA (hTR/TERC), as an additional RNA substrate for this complex. We found that depletion of either RRP6 or DGCR8 resulted in increased hTR/TERC levels, which resulted in an abnormal elongated telomere phenotype.

5.3.2 DGCR8-dependent targets of RRP6

To further define a role for the DGCR8/RRP6 complex in snoRNA degradation, we will conduct small RNA high-throughput sequencing (sRNAseq) on samples depleted of DGCR8 or RRP6 that also overexpress dominant negative versions of DGCR8 (DGCR8 dsRBD1/2) or RRP6 (D313N) at endogenous levels, respectively (discussed further throughout chapter 7). Given that the consistent but modest increase of mature snoRNA levels upon depletion of DGCR8 or RRP6 (around 1.5-2 fold) is hard to statistically detect these differences by small RNA sequencing, we deem that this approach will increase the sensitivity of this assay. Moreover, in order to correctly measure mature snoRNA levels, sRNAseq libraries should be created from RNAs <200nt; a size range routinely ignored in most sRNAseq experiments. Finally, these experiments will allow us to study in more detail the effects of DGCR8 and RRP6 binding to snoRNAs but also to tRNAs.

5.3.3 DGCR8-mediated regulation of hTR/TERC

Interestingly, the DGCR8 bound region of the telomerase RNA contains a box H/ACA like domain, which is specific for vertebrates (Mitchell et al., 1999), and has been shown to be important for biogenesis, function and localisation (Jády et al., 2004; Mitchell et al., 1999; Pogacić et al., 2000; Theimer et al., 2007). Further functional investigation identified hTR/TERC levels are directly regulated by the DGCR8/RRP6 complex, in a Microprocessor- and miRNA-independent manner, and that increases in TERC levels seen in *Dgcr8*-null mESCs correlated with an elongated telomere phenotype. Altogether this suggests the box H/ACA domain, bound by DGCR8, is also an important factor for the degradation of the hTR.

The telomerase RNP requires both TERT and hTR components to constitute an active enzyme. Therefore, it is logical to conclude that independent increases in TERT or hTR levels will not affect telomerase enzyme activities. Accordingly, only co-overexpression of both hTR and TERT dramatically increased telomerase activity. However, overexpression of the individual components have also demonstrated a 1.5 fold increase in telomerase activity *in vivo* (Cristofari and Lingner, 2006). This moderate increase in telomerase activity, could in principle be biologically relevant, since; erroneously elongation of telomeres will protect the cells from senescence. As such, the increased life span will also increase the mutational burden to a cell which could lead to malignant transformation. Accordingly, hTR has been reported to be higher in cancer cell lines (Avilion et al., 1996).

Finally, it will be of interest to investigate how relevant is the telomere phenotype in patients with the 22.11.2q microdeletion syndrome, DiGeorge syndrome, where DGCR8 is hemizygously deleted (discussed further within chapter 7).

Chapter 6. The RNA binding landscape of RRP6

6.1 INTRODUCTION

The majority of ncRNA transcripts are transcribed as long primary RNAs that require multiple processing steps by endo- and/or exonucleases before forming a mature product. The major 3'→5' decay mechanism within cells is the exosome complex which functions to degrade RNA transcripts within nucleus and cytoplasm. A number of cofactors and accessory proteins interact directly with the exosome complex, such as the NEXT, TRAMP and CBC. These complexes possess RNA binding, helicase and ribonucleotidyl transferases activities, which confer RNA substrate specificity and aid exosome-mediated degradation, thereby allowing the exosome complex to turnover and process a plethora of RNA species, and consequently regulate a number of cellular processes. In addition, the exosome core acts as a ubiquitous scaffold and associates with different complements of ribonucleases in different cellular compartments. Importantly, the only ribonuclease consistently present within all cellular compartments is RRP6.

In the previous chapter, motivated by our search for the RNA targets of the novel DGCR8/RRP6 complex, we identified direct RNA targets of RRP6 by iCLIP. Within this chapter we explore the RNA binding landscape of the RRP6-associated exosome complex in order to identify RNA substrates which require a RRP6-dependent processing step for their maturation or decay. This analysis will cover all RRP6 targets, irrespective of any overlap with DGCR8. Thus, this will illustrate a functional role of RRP6 in human cells.

6.2 RESULTS

6.2.1 RRP6 binds both coding and non-coding RNA

The RRP6 iCLIP experiments, covering overexpressed RRP6 WT and CAT, as well as endogenous protein (ENDO), produced similar RNA binding profiles. Over half of significant RRP6 clusters for each of the CLIP experiments were present within ncRNA compared to 44%-46% within coding RNA (Fig. 32 A). Interestingly, the proportion of RRP6 ncRNA binding increased to 70% when only selecting the overlapping clusters present within all three experiments (Fig. 32 A; see ALL). When these analyses were applied to gene structures, ENDO, WT and CAT experiments displayed subtly different distributions of clusters. The highest proportion of CLIP clusters were found to be present within intronic regions (ENDO, 58%; CAT, 74%; WT, 71%) and this was common between all experiments. The next most represented binding site was the 3'UTR (ENDO, 16%; CAT, 12%; WT, 12%) and coding-sequence (CDS) (ENDO, 17%; CAT, 9%; WT, 10%), whereas 5'UTR and exon-intron boundaries were the least represented amongst all experiments (Fig. 32 B). Intersection of all three CLIP experiments demonstrated reproducible binding to 3'UTR (35%) and intronic regions (30%), whereas CLIP clusters were distributed equally amongst other features (CDS, 14%; 5'UTR, 11%; exon-intron Boundaries 10%)(Fig. 32 B; see ALL). All of which indicates that RRP6 binding is enriched at the 3'UTR regions or within intronic sequences, in agreement with its known biological function as an exonuclease with 3'→5' directionality. Moreover, we showed that RRP6 can bind both coding and ncRNA, which is in line with the nuclear and cytoplasmic function of RRP6.

6.2.2 Coding RNAs

6.2.2.1 Histone mRNAs are bound by RRP6 within their 5' and 3' UTRs

A large percentage of significant RRP6 clusters were mapped within protein coding genes. A subgroup of these were histone mRNAs, which have been shown to be degraded in a RRP6-dependent manner at the end of S phase and during inhibition of DNA

replication (Graves and Marzluff, 1984; Mullen and Marzluff, 2008; Slevin et al., 2014). Our analyses of RRP6 significant clusters identified 42 histone mRNAs are bound by RRP6 in at least 1 CLIP experiment, whereas 22 histone mRNAs were bound by all 3 RRP6 CLIP experiments (Fig. 33 B). We next specifically looked at the read density of RRP6 binding sites on a particular histone mRNA, HIST1H2BJ. Examination of RRP6-ENDO CLIP read density over this transcript indicated a continuous profile of RRP6 binding, suggestive of active degradation throughout the body of the gene, with higher read densities over the 3'UTR and 5'UTR (Fig. 33 C). To examine read distribution across histone mRNAs on a global scale, we first identified expressed histone mRNAs, with at least 30% coverage in a total RNA-seq dataset (see methods for details). Histone mRNAs were then split into 20 bins and average read density was calculated over the body of the RNA. These analyses showed an enrichment of reads from RRP6 ENDO CLIP near the 3' end histone mRNAs, which was also very similar when RRP6 WT and CAT CLIPs were analysed (Fig. 33 D). The binding of RRP6 within the 3'UTR is in agreement with previous reports, suggesting a role for RRP6-dependent degradation of histone mRNAs after oligouridylation specifically within the conserved hairpin located in the 3' UTR (Fig. 33 A & C; see hairpin)(Mullen and Marzluff, 2008; Slevin et al., 2014). Unexpectedly, our analyses also showed an enrichment of RRP6 binding towards the 5'UTR of histone mRNAs, which could be indicative of short truncated histone mRNA transcripts being rapidly turned over when not needed (Fig. 33 C & D).

6.2.3 Non-coding RNAs

6.2.3.1 RRP6 is enriched in a number of small ncRNA species

To obtain the ncRNA binding profiles of RRP6, we selected those significant CLIP clusters falling within ncRNA and counted their biotypes. Again, we found similar distributions of ncRNAs between all iCLIP experiments. As expected, 50% substrates bound by RRP6 were processed transcripts (Fig. 34), indicating an extensive role for RRP6 within processing. Specifically, processed transcripts annotated by UCSC, are genes without any annotated open reading frame and undergo processing or cleavage

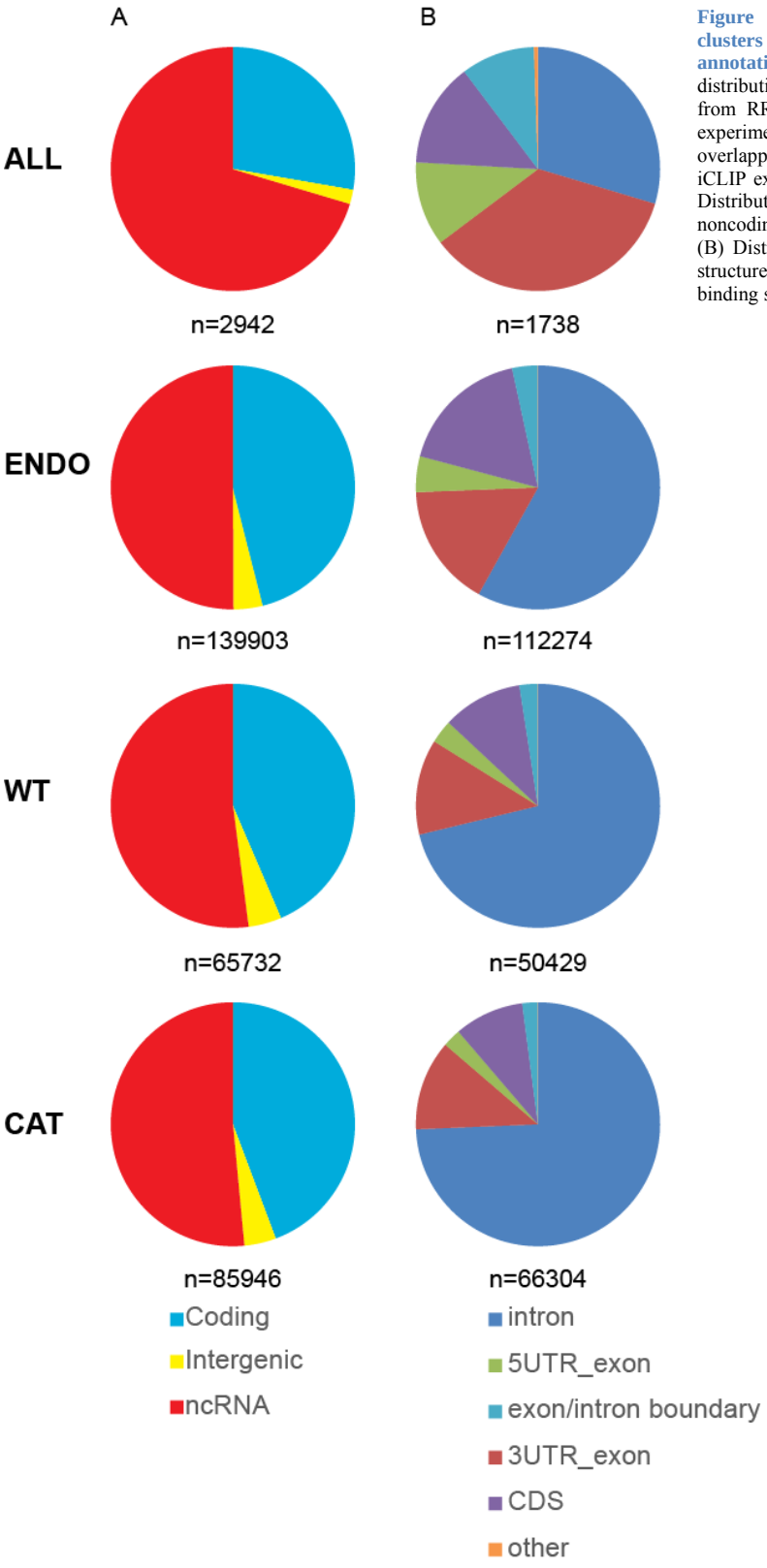


Figure 32 | Distribution of significant clusters within gene structures and ncRNA annotations. (A,B) Pie charts depicting distribution of significant RRP6 binding sites from RRP6-ENDO, -WT and –CAT iCLIP experiments. Highly reproducible, overlapping significant clusters present in all 3 iCLIP experiments are displayed as ALL.(A) Distribution of clusters throughout coding, noncoding or intergenic (unannotated) regions. (B) Distribution of clusters throughout gene structures. n= the number of significant binding sites included in analysis

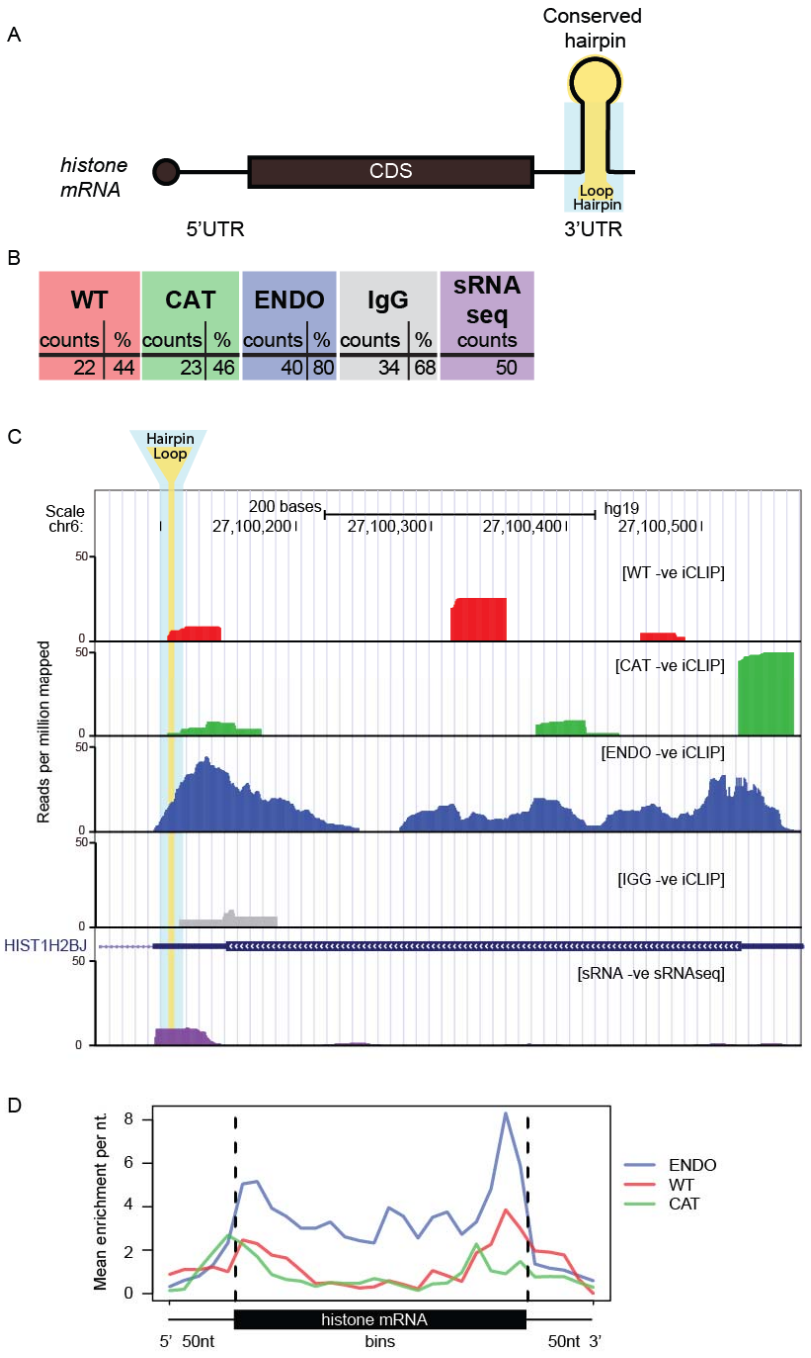


Figure 33 | RRP6 binding is distributed through histone mRNAs. (A) Schematic of histone mRNA structure. The Conserved 3'UTR hairpins loop (yellow) and stem (blue) are highlighted (B) Tables showing number of histone mRNAs bound by each iCLIP experiment. sRNAseq is a small RNA seq dataset generated from HEK293Ts (Kishore et al., 2013), used to identify expressed histone mRNAs. Percentage of expressed histone mRNAs in HEK293Tcells that are bound is shown in (%) column. (C) Genome track browser displaying read density of all iCLIP experiments, in addition to IGG and sRNA seq data set over HIST1H2BJ histone mRNA. Blue box denote conserved hairpin within 3' UTR, yellow line highlights the hairpins loop. Histone mRNA observed is on the negative strand and this indicated by arrows within histone transcript. (D) Average read density of all iCLIP experiments read density over all histone mRNAs. Histone mRNA is split into 50 bins, upstream and downstream sequences are 50nt in length and are split into 5 bins.

during their maturation. These were followed by 20% of the binding sites corresponding to antisense RNAs, plus 20% corresponding to long intergenic non-coding RNAs (lincRNAs); which is in agreement with RRP6 involvement in turnover of these ncRNAs. LincRNAs are RNA transcripts >200nt with a similar gene structure to mRNAs, but have no open reading frame (ORF). Of particular interest, were a number of small ncRNAs biotypes bound by all iCLIP experiments. These were snoRNAs, tRNAs, snRNAs and miRNAs, all of which have been previously identified as RRP6 substrates for processing or degradation (Fig. 34). Altogether, these data indicate RRP6 can bind a number of small ncRNAs.

6.2.3.2 Comparative CLIP analyses identify RRP6 binding hotspots within 45S rDNA clusters

Within humans, 18S, 5.8S and 28S rRNA are transcribed as one long *cis* transcript; termed 45S rRNA. This transcripts then goes through a series of multiple endo- and exoribonucleolytic processing steps that are required to release the mature rRNAs. The best characterised substrate of RRP6 so far in humans and yeast is the 5.8S pre-rRNA. However, the direct RNA interactions of RRP6 within pre-rRNA have not been thoroughly investigated within humans. For this reason, we next analysed CLIP density over the 45S rDNA cluster. However, due to the repetitive nature of this cluster within the human genome, iCLIP reads mapping to this region are likely to be disqualified from

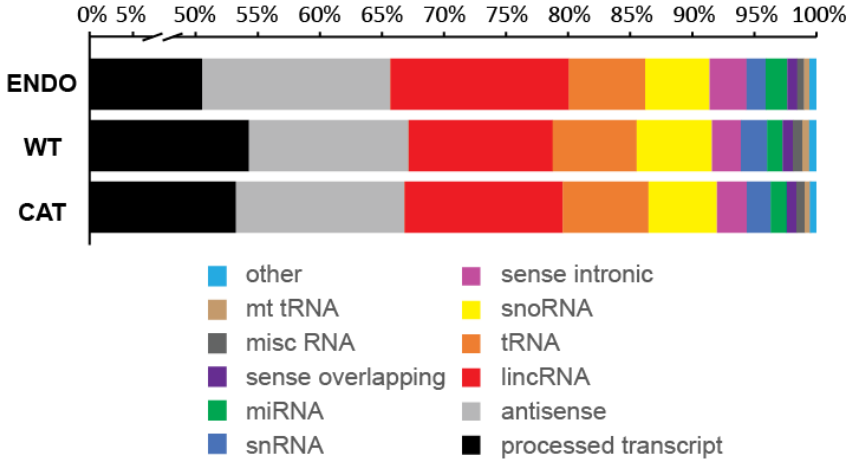


Figure 34 | Distribution of significant clusters RRP6 CLIP clusters throughout ncRNAs. Histogram displaying percentage of significant clusters bound within ncRNA biotypes. Values were calculated by summing the number of different ncRNA biotypes bound, and then represented as a percentage of all clusters counted.

subsequent analyses, since they cannot be uniquely mapped. Therefore, we aligned CLIP reads from every experiment to a consensus sequence of 45S rDNA. Reads were pre-processed to ensure the reads average Phred score across a 3 base sliding window was ≥20 (see material and methods for further details). In total 519,128 and 502,034 reads from WT and CAT CLIP, respectively, were aligned to the 45S consensus sequence, which comprised 17% and 19% of their total reads. We found high read densities and coverage of CLIP tags over mature rRNA sequences (Fig. 35). Although these signals may represent actual binding, non-specific interactions with mature rRNA are common due to these transcripts accounting for ~95% of total RNA within the cells. Strikingly, density profiles within less abundant pre-rRNA, particularly ITS-2, from both CAT and WT CLIP profiles were non-overlapping. Within a 30nt region, immediately downstream of mature 5.8S rRNA, a large enrichment of CAT over WT CLIP reads was evident (~6-8 LOG2 fold) (Fig. 36). The lower density of reads present in RRP6 WT CLIP suggests efficient and rapid processing of pre-rRNA, whereas we hypothesise that RRP6 CAT is acting as a dominant negative mutant by stabilising protein:RNA interactions and blocking processing.

Interestingly, this was not the only region in pre-rRNA with this difference in CLIP read densities. The ITS-1 also contained a major enrichment of CAT CLIP read density over WT (Fig. 37) in a 30nt window, which is reminiscent to that seen downstream of 5.8S pre-rRNA (Fig. 36).

ITS-1 also undergoes endonucleolytic cleavages at conserved sites in order to liberate mature 18S and 5.8S rRNA from the longer pre-rRNA transcript. Interestingly, RRP6 binding is not witnessed throughout ITS-1 and there are regions depleted of CLIP signal (Fig. 37; see arrows). This indicates that the whole ITS-1 is not processed/degraded by RRP6. Furthermore, regions which are sharply depleted of CLIP signal may indicate the 5' end of transcripts, created by endonucleolytic cleavage, which are being degraded by RRP6. For example, absence of CLIP density was seen upstream of the 30 nt window and downstream of the 3' end of 18s rRNA (Fig. 37; see arrows).

Finally, we focused on the 5'ETS, another pre-rRNA region that had been postulated to be degraded by RRP6 (Sloan et al., 2013). Analysis of the CLIP read densities over 5'ETS, did not show clear enrichments of RRP6-CAT vs RRP6-WT CLIP tags, as was the case for previous regions (Fig. 38). However, the high read density from both RRP6 WT and CAT CLIP profiles from 421nt-2501nt within 5'ETS (Fig. 38; see red line), indicate this region may be targeted by RRP6, as previously suggested (Sloan et al., 2013). On the other hand, the absence of CLIP tags within 5'ETS from 1nt to 420nt indicates RRP6-independent turnover (Fig. 38).

Altogether these data show that the overexpression and binding of RRP6 CAT to RNA substrates stabilises protein:RNA interactions, and may act as a dominant negative mutant which results in a stabilisation of RNAs processed by RRP6. These observations raise the possibility of using RRP6 CAT in future studies, to detect and identify rapidly degraded targets. Moreover, increased RRP6 CAT CLIP density indicates specific roles for RRP6-mediate degradation and/or processing during rRNA maturation, whereas depletion or absence of CLIP signal may indicate a conserved sites for endonucleolytic cleavage within pre-rRNA.

6.2.3.3 RRP6 binds pre-snoRNA loci

SnoRNA biogenesis has been characterised in detail within yeast, where the exosome and Rrp6p are known to have an essential role in pre-snoRNA processing (Allmang et al., 1999, 2000; van Hoof et al., 2000a). In contrast, very little is understood about pre-snoRNA processing in humans. In order to study the role of RRP6 in pre-snoRNA processing, we examined the RRP6 CLIP read density over mature and neighbouring intronic sequences. We found that RRP6 is reproducibly distributed throughout the mature snoRNA sequences of SNORD18C and SNORA71B (Fig. 39 A & B). Interestingly, the read profiles from RRP6 CAT and ENDO extended beyond the 3'EAG (3' End of the Annotated Gene), into the intronic precursor snoRNA sequences; whereas these extensions were not seen upstream of the 5'SAG (5' Start of the Annotated Gene). Specifically, RRP6 CAT CLIP, but not RRP6 WT, read profiles extended ~31nt and ~26nt downstream of the 3'EAG for SNORD18C and SNORA71B, respectively

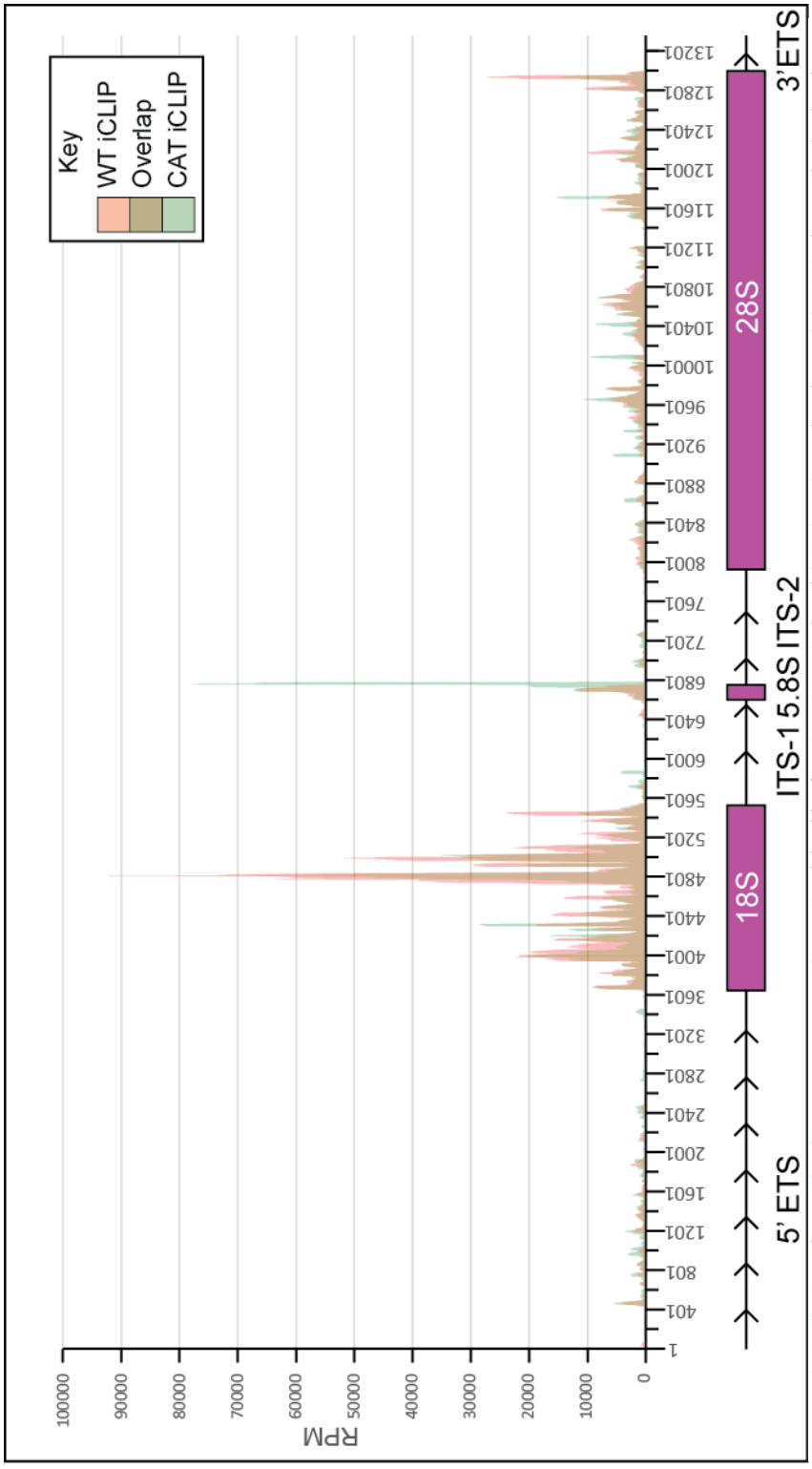


Figure 35 | Distribution of RRP6 binding over 45S rRNA cluster. Read densities from RRP6 WT (red) and CAT (green) are plotted over 45S rDNA cluster. The density plots are overlaid with each other, therefore dark regions indicate overlap of binding profiles. Datasets were normalised to library size and displayed as reads per million (RPM). (A) A schematic of 45S rDNA cluster and the positions of 18S, 5.8S and 28S rRNAs are indicated in purple. The 5' and 3' External transcribed spacers (5'ETS and 3'ETS) and intervening internal transcribed spacers 1 and 2 (ITS-1 and ITS-2).

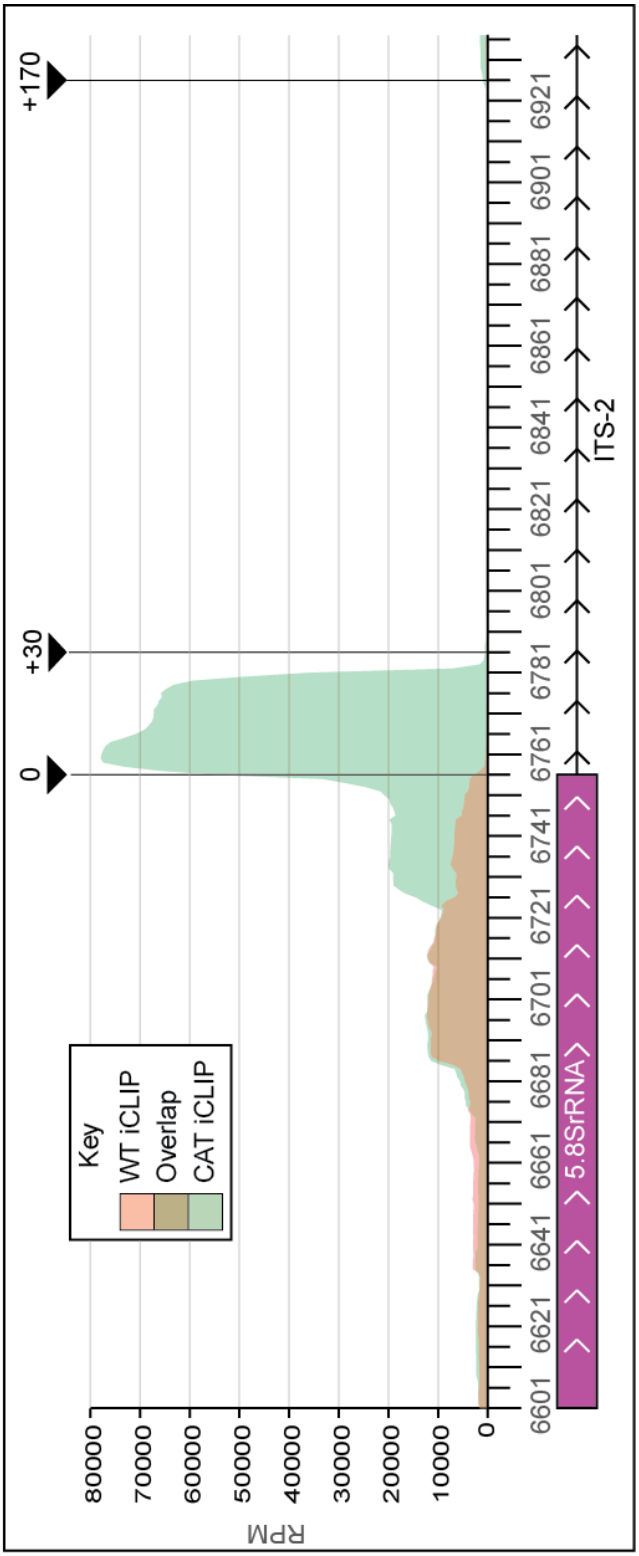


Figure 36 | RRP6 binds 5.8s pre-rRNA. The region displayed is a zoomed in from Fig. 35. It contains 5.8S rRNA and the downstream ITS-2. Arrows indicate distance from 3' end of 5.8s rRNA.

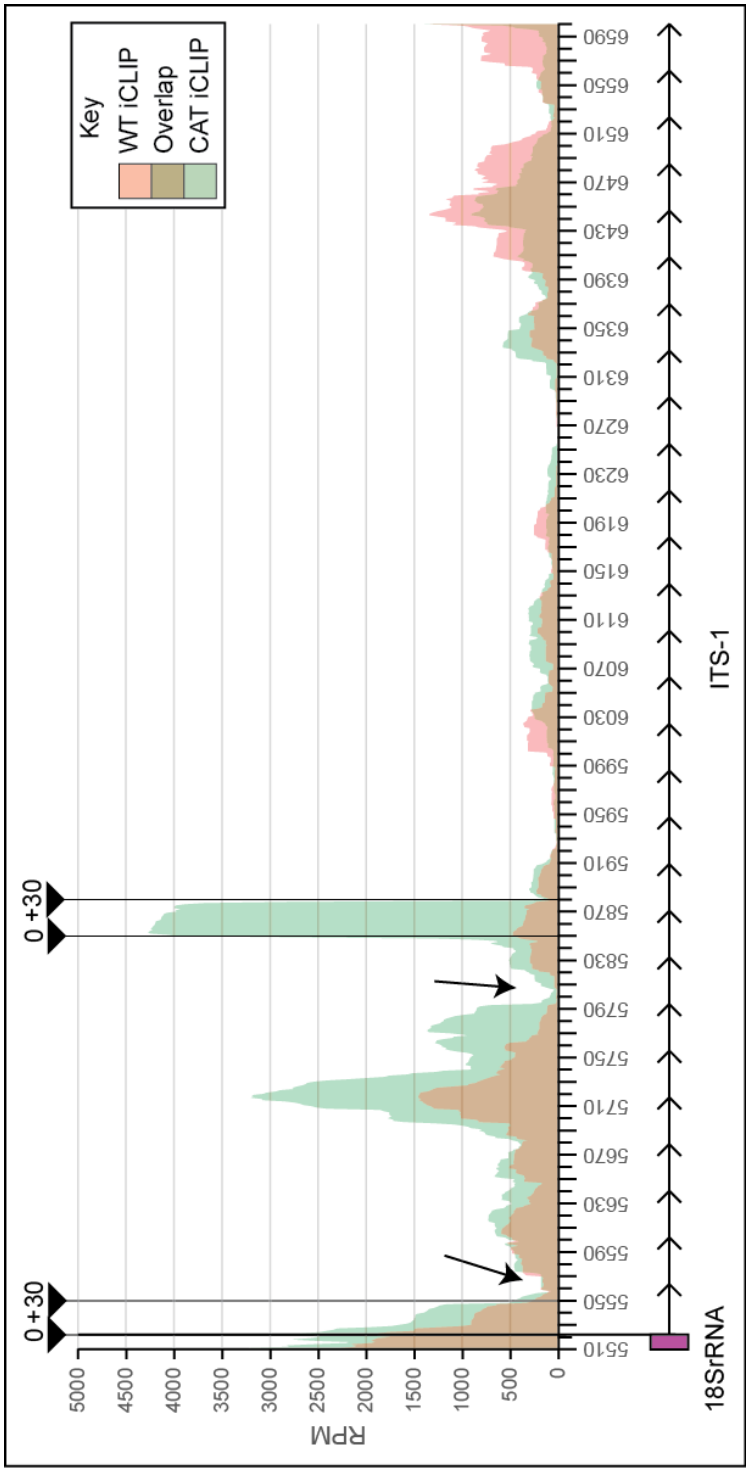


Figure 37 | RRP6 binds ITS-1. Zoomed in region from Fig. 35 displaying 18S and ITS-1, arrows identify depletion in read density, whereas horizontal lines depict 30nt periodicity.

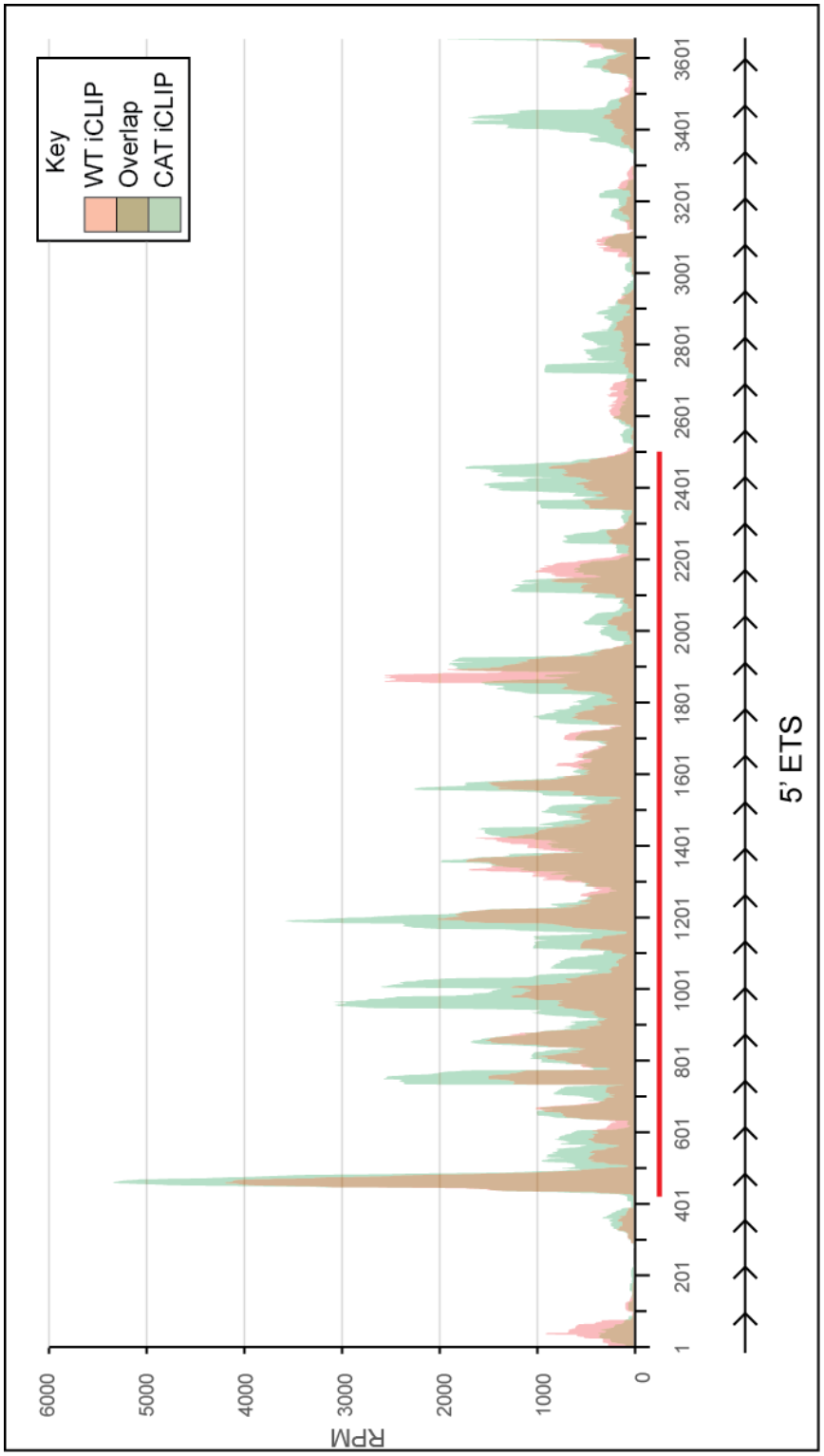


Figure 38 | RRP6 binds 5'ETS. Zoomed in region from Fig. 35, displaying 5'-ETS. Red horizontal line indicates high CLIP density over fragment of 5'ETS.

(Fig. 39 A & B). In addition, RRP6 CAT stabilised these extended forms, exhibiting at least 2.5 LOG2 increase in read density, when compared with RRP6 WT, for both the snoRNAs shown (Fig. 39 A & B). These data suggest that RRP6 CAT is binding and stabilising 3' flanking regions of snoRNAs and therefore may be involved the removal of 3' intronic sequences from pre-snoRNAs. In addition, the stabilised RNAs present in RRP6 CAT CLIP are not found in the sRNAseq dataset (Fig. 39 A & B; see green and red arrows), indicating these pre-snoRNAs are of low abundance in transcriptomic studies.

Next, in order to establish if RRP6 is involved pre-snoRNAs processing globally, we quantified CLIP read densities over snoRNA loci at defined genomic intervals in order to detect RRP6 binding to pre-snoRNAs sequences. These intervals were located over the mature snoRNAs, 50nt upstream of the 5'SAG (5'flank) and 50nt downstream of the 3'EAG (3'flank). Specifically, reads mapped within mature snoRNAs, and not overlapping flanking regions were classified as mature reads, whereas reads overlapping with 5' and 3' flanks were assigned to their respective category (for schematic see Fig. 40 A). In parallel, we conducted the same analysis on an RNAseq dataset obtained from 20nt – 200nt RNA species isolated from HEK293Ts (sRNAseq), hence, containing mature full length snoRNAs (60nt – 150nt) and other mature sRNA species (Kishore et al., 2013), in order to understand the distribution of sRNAseq reads associated with snoRNAs.

The analyses of the sRNAseq in HEK293Ts revealed that 20% of the reads mapped to 294 mature snoRNA loci. From these reads, 85% accounted for SNORD; whereas 13% and 2% accounted for SNORA and SCARNA subtypes, respectively (Fig. 40 B). This suggests SNORDs are the most abundant class of snoRNAs within HEK293T cells. In addition, we calculated the distribution of reads mapping to 5'flanks, mature or 3' flanks using the strategy depicted in Figure 40 A for the sRNAseq, and found that 65% of reads associated with snoRNA loci map exclusively to the mature sequence, whereas 16% and 18% are distributed between their 5' and 3' flanks, respectively (Fig. 40 D; see sRNAseq).

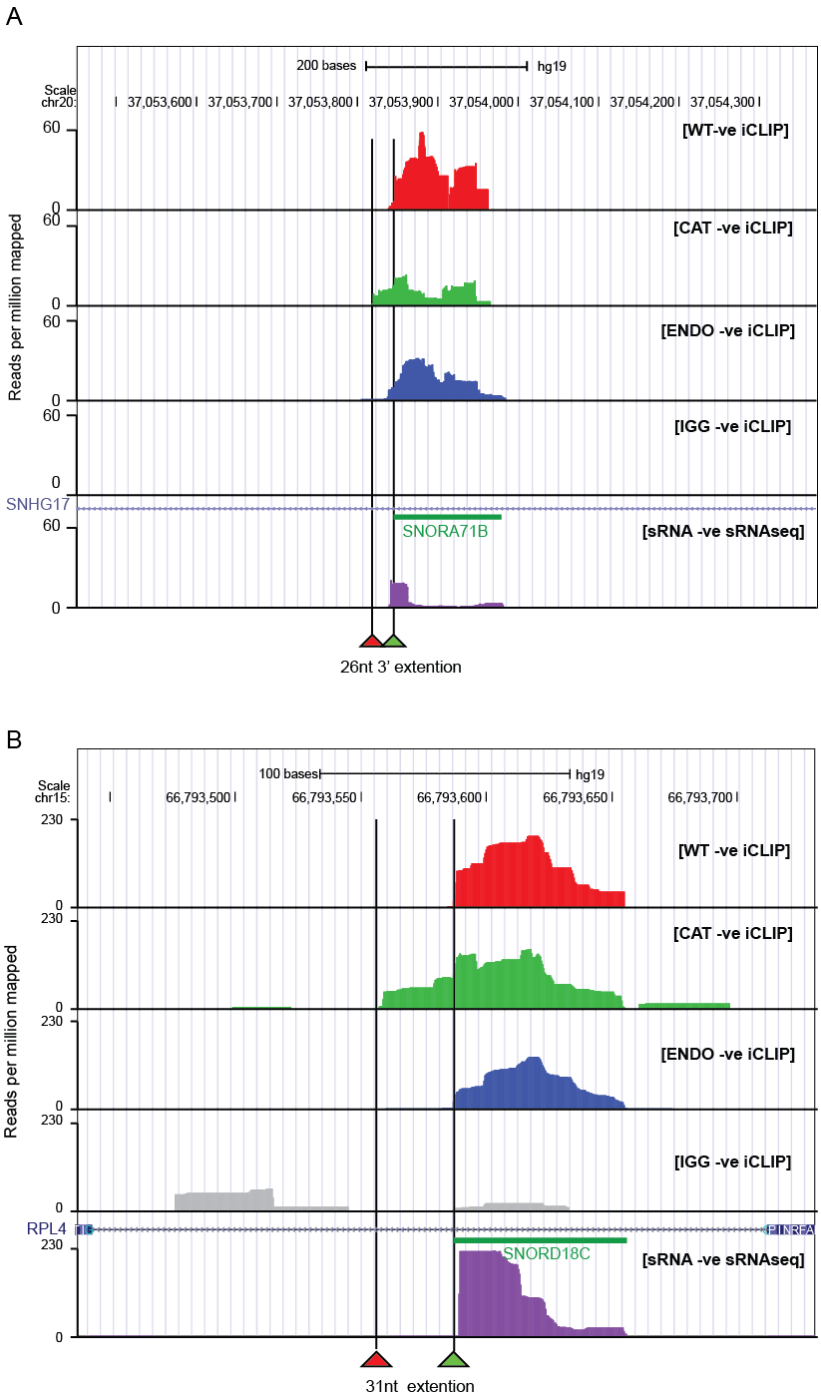
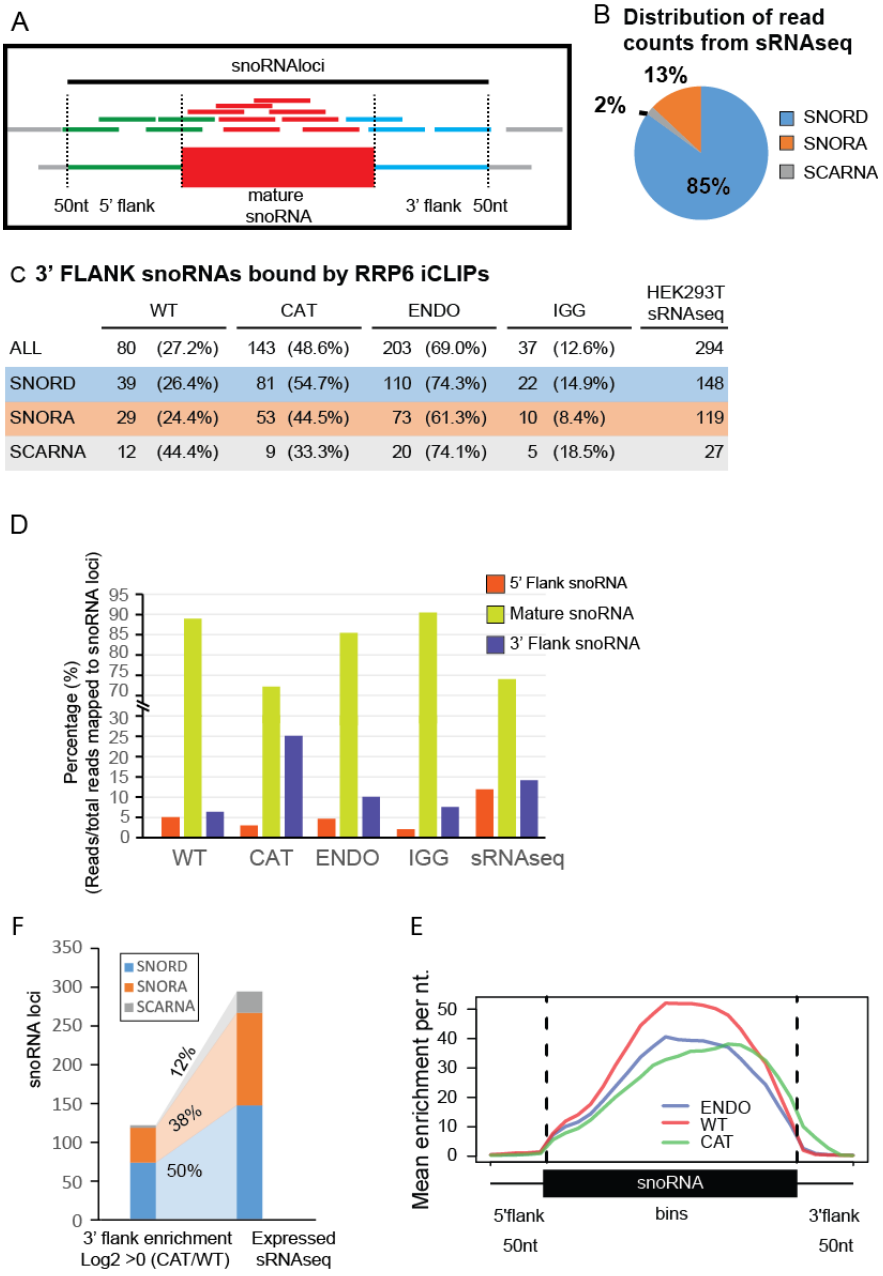


Figure 39 | RRP6 CAT stabilises pre-snoRNAs. (A,B) Genome track browser displaying RRP6-WT (red); -CAT (green); -ENDO(blue) or IGG(grey) iCLIP read densities plotted over snoRNA loci. Also a sRNAseq (purple) is displayed for comparison (Kishore et al., 2013). (A) Displaying SNORA71B within SHG17. (B) Displaying SNORD18C which is located in RPL4. Green arrows annotate the 3'EAG (End of the Annotated Gene), whereas red arrow indicates length 3'extended snoRNAs bound by RRP6 CAT. In each case, the length of pre-snoRNA stabilised is plotted below the arrows.

Our data showed that RRP6 ENDO bound 69%, whereas RRP6 WT and CAT, respectively bound 27% and 49% of the 294 expressed snoRNAs within their 3'flanking region (Fig. 40 C). All CLIP experiments detected all snoRNA classes, however, RRP6 WT was biased towards an enrichment of binding to SCARNAs (SNORD, 26%; SNORA,24%; SCARNA, 44%), whilst RRP6 CAT favourably bound SNORA and SNORD classes (SNORA,45%; SNORD,54%: SCARNA,33%); whereas RRP6 ENDO had no obvious preferences (SNORD,74%; SNORA, 61%; SCARNA, 74%) (Fig. 40 C). To identify if specific regions of snoRNA loci were enriched within our CLIP datasets, we again calculated the distribution of sRNAseq and CLIP reads associated with the 5'flank, mature and 3'flank genomic intervals over snoRNA loci. From these analyses we identified a higher frequency of CAT CLIP reads within 3'flanking regions of snoRNAs when compared with RRP6 ENDO or RRP6 WT CLIP (Fig. 40 D). This indicated RRP6 CAT is specifically bound to 3' flanks of snoRNA loci. We selected all 3' flanking regions of snoRNA loci bound by at least one CLIP experiment, and calculated the read density ratio between RRP6 CAT and WT CLIP dataset. This demonstrated that 50% of all 3'extended snoRNAs bound by RRP6 within our CLIP experiments were stabilised specifically by RRP6 CAT (>0 LOG2 CAT/WT read density) (Fig. 40 E). This accounted for 50% of all SNORD snoRNAs expressed in HEK293T cells; in addition to 38% and 12% of all SNORA and SCARNA snoRNAs expressed in HEK293Ts (Fig. 40 E). In agreement with this, average read profiles of all CLIP experiments over expressed snoRNA loci also highlighted an enrichment of CLIP density within the 3' flanking region of snoRNAs (Fig. 40 F). Conversely, the distribution of read densities associated with RRP6 WT and ENDO are lower than sRNAseq reads densities over 3'flanks of snoRNA loci, but higher over mature loci (Fig. 40 D). Altogether, these data suggest that again overexpression of RRP6 CAT acts like a dominant negative, stabilising snoRNA processing intermediates.

6.2.3.4 20-30nt extended forms of snoRNAs are stabilised by RRP6 CAT

To better understand the role of RRP6 within pre-snoRNA processing, we next identified the length of 3' extended snoRNAs bound by RRP6 on a global scale. To



C 3' FLANK snoRNAs bound by RRP6 iCLIPs

	WT	CAT	ENDO	IGG	HEK293T sRNAseq
ALL	80 (27.2%)	143 (48.6%)	203 (69.0%)	37 (12.6%)	294
SNORD	39 (26.4%)	81 (54.7%)	110 (74.3%)	22 (14.9%)	148
SNORA	29 (24.4%)	53 (44.5%)	73 (61.3%)	10 (8.4%)	119
SCARNA	12 (44.4%)	9 (33.3%)	20 (74.1%)	5 (18.5%)	27

Figure 40 | Genome wide analysis of RRP6 binding to snoRNAs. (A) Schematic of strategy used to identify iCLIP reads within 5' flank, mature and 3' flank. (B) Abundance of expressed snoRNAs within HEK293T cells (Kishore et al., 2013). (C) Table displaying the number of snoRNAs bound within their 3' flanks by each RRP6 iCLIP experiment using strategy depicted in (A). Percentages shown are relative to total number of snoRNAs, within each subtype, that are expressed in HEK293Ts. (D) Distribution of reads mapping to different regions of snoRNA loci, calculated using strategy depicted in (A), and then normalised to total number of reads mapped to snoRNA loci within each experiment. (E) Histogram displaying all snoRNAs with high read density in RRP6 CAT CLIP density over RRP6 WT CLIP (>0 LOG2) within their 3' flank region. Total number of expressed snoRNAs in HEK293T sRNAseq dataset indicated by "Expressed sRNAseq". Percentage of expressed snoRNAs within HEK293Ts that display enrichment of RRP6 CAT CLIP density when compared to RRP6 WT CLIP density (>0 LOG2) within their 3'flank. (F) Average read density profile of all RRP6 iCLIP experiments plotted over all snoRNAs expressed within HEK293T cells. The mature sequence is split into 20 bins, while 5' and 3' flanks are represent by 50nt upstream and downstream stretches adjacent to annotated snoRNA; these regions are split into 5 bins each.

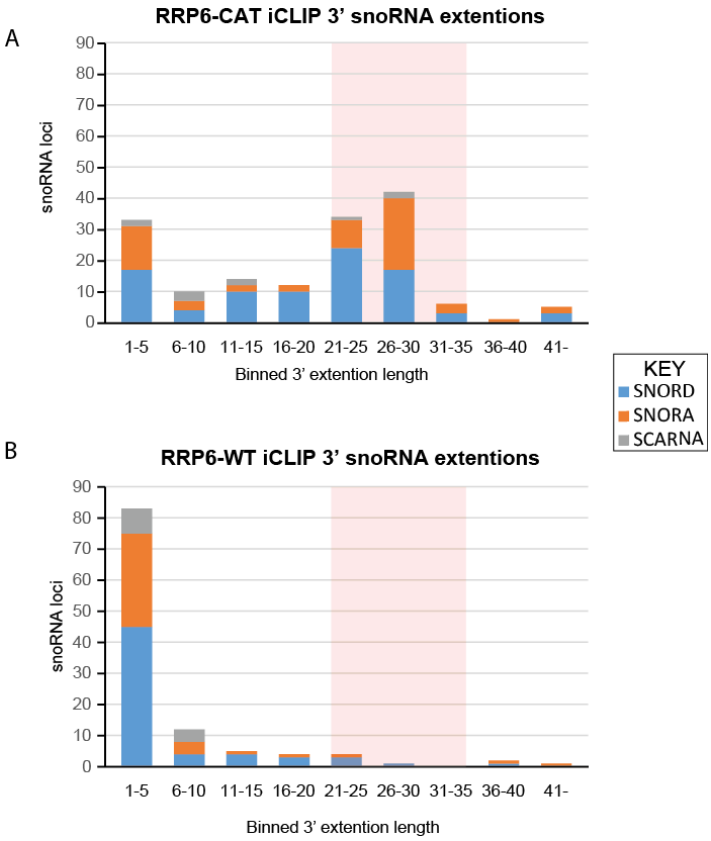


Figure 41 | RRP6 binds to 3' flank of pre-snoRNAs. (A,B) Distribution of the length of pre-snoRNAs bound by RRP6 CAT (A) and WT (B) iCLIP. Red box highlights binned values between 21-30nt.

do so, we calculated the maximal distance of mapped reads within the 3'flank of snoRNA loci from the 3' end of mature snoRNAs (for example, see red and green arrows in Fig. 39 A & B). Interestingly the maximal distance between the 3' end of the mapped read and the snoRNA 3'EAG varied between RRP6 WT and CAT CLIP datasets. Specifically, RRP6 CAT CLIP read profiles extended between 1nt-5nt and 21nt-30nt, whereas RRP6 WT profiles extended 1nt-5nt beyond snoRNAs 3'EAG (compare red boxes in Fig. 41 A & B). This finding that RRP6 CAT stabilised snoRNA + 30nt is in agreement with RRP6 CAT read density profiles of SNORA71C and SNORD18C and also 5.8S rRNA processing (compare red boxes in Fig. 41 A & B with 30nt window downstream of 5.8S rRNA in Fig. 36). Altogether these data illustrate that RRP6-mediate processing of pre-snoRNAs is required for the resection of the terminal 30nt downstream of 3' end of the mature snoRNA sequence.

A

	Wt		CAT		ENDO		IgG		sRNAseq
	Counts	%	Counts	%	Counts	%	Counts	%	Counts
tRNA loci bound	303	63.12	268	55.83	325	67.70	81	16.87	480

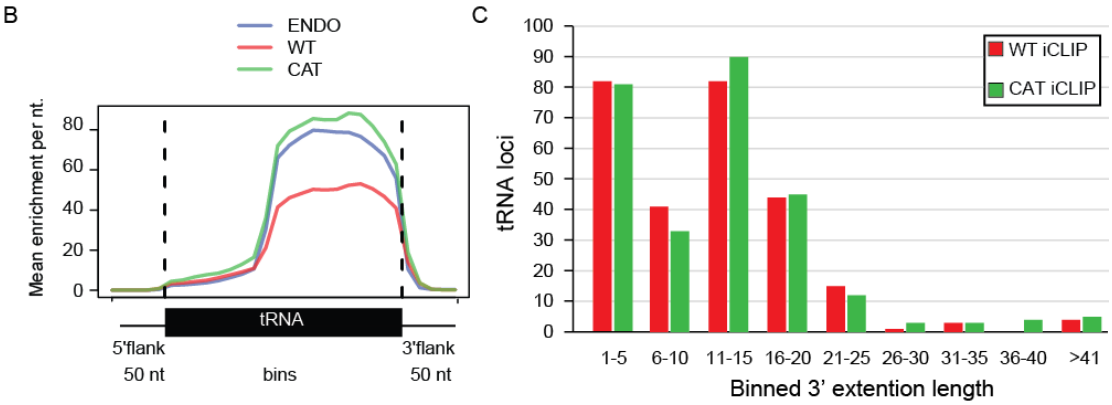


Figure 42 | RRP6 is bound to the 3' end of tRNAs. (A) Number of tRNA loci bound by each individual RRP6 CLIP experiments and IGG. Percentage or expressed tRNAs bound by RRP6 is shown. (B) Average read profile of all CLIP experiments over tRNA loci. The mature sequence is split into 20 bins, while 5' and 3' flanks are represent by 50nt upstream and downstream stretches adjacent to annotated tRNA; these regions are split into 5 bins each. (C) Genome wide analysis of the length of tRNA 3' flanks bound by RRP6 in WT and CAT CLIP.

6.2.3.5 tRNAs are bound by RRP6

Analysis of significant CLIP clusters identified tRNAs as another sRNA bound by RRP6. To identify the context of RRP6 binding within tRNAs, we first identified the number of tRNA loci bound by all RRP6 iCLIP experiments. This analysis revealed that RRP6 ENDO bound 68%; WT bound 63%; and CAT count 55% of tRNAs expressed within HEK293T cells, respectively (Fig. 42 A). Next, to identify the coverage of binding over tRNA globally, we identified the average binding profile for all RRP6 iCLIP experiments over tRNA loci, using the same approach as for Figure 40 E. This identified that RRP6 CAT and ENDO CLIPs were enriched in mature tRNAs more than RRP6 WT, whereas all CLIP experiments bound to 3' flanking regions of tRNAs to a similar degree (Fig. 42 B). Next we identified the maximal length profiles of the 3'extended pre-tRNAs, using the same approach as detailed for pre-snoRNA analysis in Figure 41. The profiles generated from these analyses were different to pre-snoRNAs. RRP6 bound relatively few fragments between 21-30nt and a clear biphasic pattern was observed, one ranging from 1nt to 10nt and a second ranging from 11nt to20nt) for both RRP6 WT and CAT CLIP

experiments (Fig. 42 C). Interestingly, closer examination of the 3' end extension bound by RRP6 identified that these extensions coincided with the RNAPIII terminator sequences, which is a T rich consensus (Fig. 43 A & B; see red boxes).

Altogether, these data suggest that RRP6 is involved in the turnover of mature snoRNAs and tRNAs given that we can detect binding to the mature sequences. Additionally, RRP6 can also participate in the 3' end processing of pre-snoRNAs and pre-tRNAs.

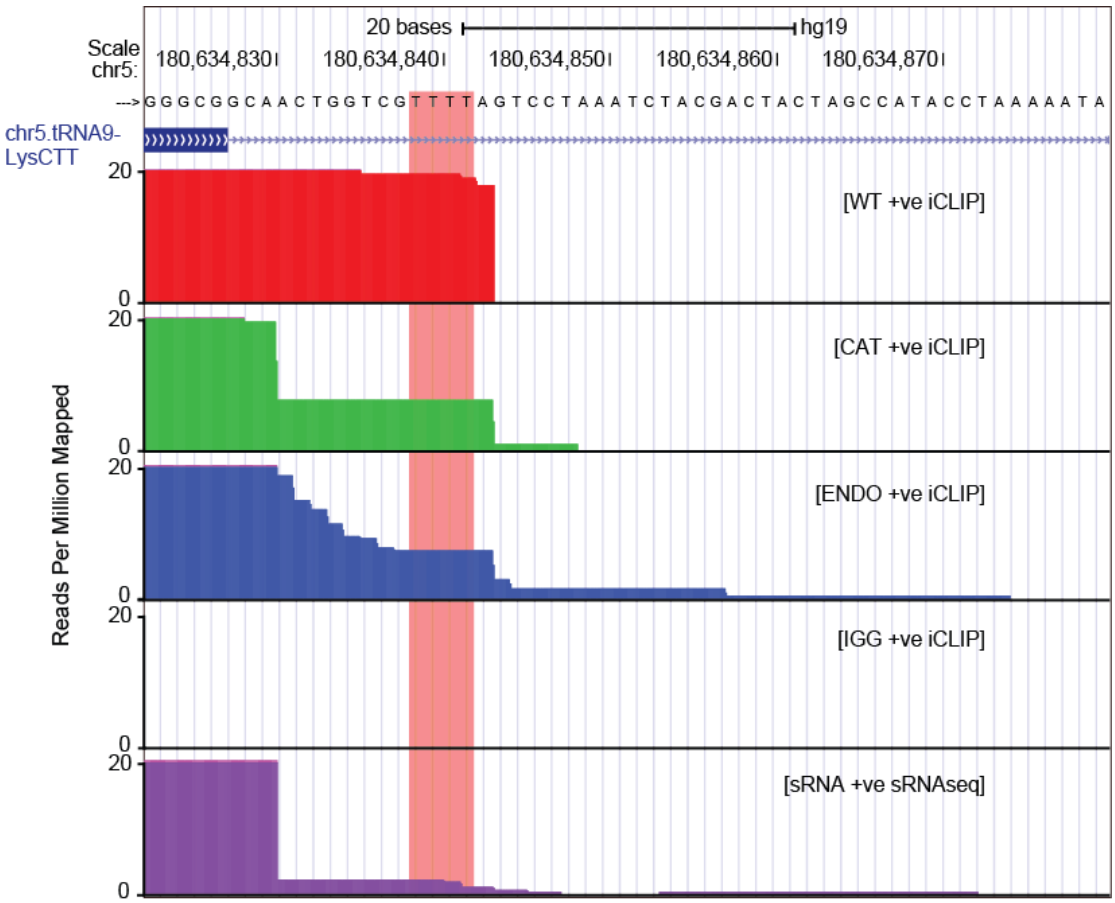


Figure 43 | RRP6 binds to 3'flanks of tRNAs. RRP6 iCLIP and sRNAseq read density plot over tRNA9-LysCTT. The scales have been restricted to 20 RPMM in order to visualise densities over 3' flank regions. Red boxes highlight T rich consensus important for transcriptional termination of RNAPIII.

6.3 DISCUSSION

6.3.1 Summary

Initially, the search for DGCR8/RRP6 substrates led to us establishing the RNA binding landscape of RRP6 protein. For this purpose, we took a two-step approach, purifying protein:RNA complexes from endogenous and overexpressed RRP6 (functional) and overexpressed catalytically dead RRP6 (non-functional). Comparative analysis of RNA binding profiles from these iCLIP datasets identified reproducible binding within similar groups of coding and non-coding RNA species, and of particular interest was the observed binding to histone mRNAs, rRNA, snoRNAs, and tRNAs, which we decided to investigate further.

6.3.2 Histone mRNAs are bound by the RRP6 associated exosome

Histone mRNA transcription is tightly regulated throughout the cell cycle. An upregulation of histone mRNA transcription during G1/S phase is required to boost histone protein levels in order to package newly synthesised DNA. This is followed by rapid histone mRNA turnover at end of S phase after DNA synthesis is complete, which prevents further Histone protein translation (Graves and Marzluff, 1984). The turnover of histone mRNAs is initiated by Eri1, an exonuclease that degrades into the dsRNA structure of the conserved hairpin at the 3'UTR. Subsequently, oligouridylation by TUTases reprimes the 3' end of the histone mRNA for degradation by RRP6 (Mullen and Marzluff, 2008). The requirements for Eri1 to degrade into the dsRNA structure of the 3' hairpin is presumably due to the incapacity of RRP6 to exonucleolytically cleave strong RNA secondary structures (Januszyk et al., 2011). Accordingly, the observed increased CLIP read density for RRP6 within or immediately after the conserved hairpin sequence supports this mechanism (Fig. 33 B & C).

In addition, we detected RRP6 binding within 5'UTR regions of histone mRNAs, which had not been reported before. This observation could be explained by two alternative mechanisms. Firstly, the binding of RRP6 to histone mRNA 5'UTRs represents an intermediate step of the rapid cytoplasmic turnover of histone mRNA levels

at the end of S phase. This mechanism would probably involve an endonucleolytic cleavage step within the 5'UTRs to create a 3' end viable for exosome degradation. Alternatively, this could indicate that short truncated histone mRNAs are rapidly turned over in the nucleus, which would be presumably formed by premature transcriptional termination of RNAPII. In order to discriminate between these two alternative mechanisms, it will be essential to identify which portions of the histone mRNAs are stabilised after RRP6 inactivation; and secondly, determine in which cellular compartment this regulation takes place.

6.3.3 RRP6 participates in 3' end sculpting of short ncRNA species

The nuclear exosome complex can associate with DIS3 and RRP6. These ribonucleases have different processivity; DIS3 is highly processive, whereas RRP6 has distributive activity. It has been suggested these ribonucleases must work cooperatively in order to process or degrade RNAs (Allmang et al., 2000; Schneider et al., 2012). Interestingly, we detected RRP6 binding immediately downstream of 3'EAG for tRNAs, snoRNAs and 5.8S rRNA, which suggests a role of RRP6 in 3' end pre-RNA processing. Moreover, increased read densities within RRP6 CAT profiles demonstrated non-functional RRP6 was bound to pre-snoRNAs more than its functional counterparts (WT and ENDO) (Fig. 40 E). This suggests the inactive RRP6 could stabilise precursor sequences, which are normally rapidly turned over by RRP6, which is evidenced by the lower density of reads in RRP6 WT and ENDO CLIP profiles. Following these observations, we found the length of 3' extensions stabilised by RRP6 CAT differed depending on the subtype of ncRNA. Specifically, 5.8S pre-rRNA and pre-snoRNAs extensions were between 20nt-30nt downstream of the 3'EAG, whereas tRNAs exhibited two groups of 3' extended RNAs, between 1-10nt and 11nt-20nt. These observations could indicate the presence of two different mechanisms for RRP6-mediated processing.

6.3.4 The mechanisms governing 3' end processing are similar between snoRNA 3' and rRNA

Both snoRNAs and rRNA biogenesis mechanisms depend on the removal of long intronic precursor sequences by exonucleases, and binding of RNPs to their mature sequences in order to stabilise their RNA secondary structures and protect from the unspecific degradation by other nucleases. Interestingly, when analysing the binding of RRP6 CAT to these sequences, we observed a stabilisation of the downstream 30 nt in both pre-snoRNAs and 5.8S pre-rRNA. This highlighted a role for RRP6 in the removal of these sequences, and not those downstream of these 30nt. Importantly, these results are in line with previous reports in yeast that Dis3p is required for the processing of 7S to 5.8S+30nt (Allmang et al., 1999).

In addition to RRP6, the nuclear exosome associates with DIS3. Unlike RRP6, DIS3-mediated exonucleolytic degradation requires the RNA substrate to be channelled through the core, to its exonuclease active site. Interestingly, the channel within the exosome core has been demonstrated to be 30nt in length and wide enough to accommodate only unstructured ssRNA (Bonneau et al., 2009). Therefore it would be attractive to postulate the stabilisation of the 30nt fragments is due to a switch in ribonuclease activity of the exosome complex; from DIS3, core-dependent processing; to RRP6 core-independent processing. In this scenario, we envisage 3' flanking sequences from both pre-snoRNA and 5.8S pre-rRNA being threaded through the exosome core to be degraded by DIS3 until, the bulky RNPs, which protect the mature RNA sequence from degradation and are too large to fit into the exosome core, collide with the entrance to the exosome complexes. At this point, DIS3 processing would become incompatible, due to the bulky RNPs restricting further processing of the 3' flanking sequence. What remains is a 30 nt sequence extending from the 3'EAG through the core. At this point, a switch of processing from a DIS3-mediated and channel-dependent to a RRP6-mediated and channel-independent would occur. The advantage of RRP6 exonucleolysis is that there are no issue with steric hindrance cause by bulky RNPs, as substrates do not pass through

the core. Therefore the remaining 30nt would be processed by RRP6 until the bound RNPs block any further decay, thereby defining the 3' end of the ncRNA.

6.3.5 tRNA 3' end formation

In contrast with snoRNAs and rRNAs, tRNAs are transcribed by RNA polymerase III as short transcripts, initiated at their own promoters and terminated by a T-rich consensus sequence, a short distance downstream of the 3'EAG (Cozzarelli et al., 1983). Subsequent cropping by endonucleases define 5' and 3' tRNA ends, however recently it was reported RRP6 may play a minor role in 3' end processing of tRNAs (Skowronek et al., 2013). This in agreement with the observation that RRP6 bind 3' extended forms of tRNAs.

In our experiments, the periodicity of pre-tRNA fragments stabilised by RRP6 CAT was shorter than pre-5.8S rRNA and pre-snoRNAs, suggesting that channel-dependent processing is not required. Biphasic groups of extended pre-tRNAs between 1-10nt and 11nt-20nt were stabilised in both RRP6 CAT and WT experiments. Interestingly, for at least the examples shown, these two groups could be identifying the maximal distance of the RNAPIII termination signal from the 3' end of the gene (Fig. 42 A). Further genome wide analyses of terminator sequence proximity to the 3'EAG of pre-tRNAs must be performed in order to come to this conclusion.

Chapter 7. Discussion & Future work

7.1 DISCUSSION

7.1.1 DGCR8-mediated RRP6-dependent snoRNA degradation

The primary aim of this thesis was designed to identify the putative ribonuclease partnering DGCR8 to degrade snoRNAs. Biochemical analysis of DGCR8 complexes identified DGCR8 interacted with the exosome component RRP6. These results led us to globally characterise the RNA targets of RRP6, in order to identify overlapping targets with previous DGCR8 HITS-CLIP data (Macias et al., 2012).

7.1.2 Evolution of DGCR8 as a dsRNA binding adaptor

There is no DGCR8 yeast homologue, and given the requirement for a wider range of ancillary proteins to deal with increased RNA complexity in higher order eukaryotes, numerous RNA binding proteins maybe required to deal with RNA turnover within the cell. However, Rnt1p, a yeast RNase III enzyme, has been shown to perform similar functions to the Microprocessor complex, such as premature transcriptional termination (Dhir et al., 2015; Rondón et al., 2009). Rnt1p contains a single dsRBD domain and a single RNase III domain. The dsRBD domain binds ss-dsRNA, and directs the RNase III domain to cleave bound substrates. Similarly, within humans, the dsRNA binding component of the Microprocessor, DGCR8, guides the RNase III enzyme, Drosha, to cleave pri-miRNA-hairpin structures. Essentially, DGCR8 and Drosha are performing functions analogous to the dsRBD and RNaseIII domains of Rnt1p, except that they are two distinct proteins. This poses an interesting question why the Microprocessor complex exists as a modular complex, whereas another RNase III enzyme, Dicer, has retained dsRNA binding and endonucleolytic activity. This alone may be suggestive of DGCR8 and Drosha being able to form functional complexes with other proteins.

Recently, experiments by Roth et. al suggest DGCR8 has the potential to bind a wide array of RNAs with different structures, and that target specificity may arise from its canonical partner within the Microprocessor complex, Drosha (Roth et al., 2013). If this holds true, DGCR8 decoupled from Drosha, could bind a different set of cellular

RNAs. Interestingly, DGCR8 was shown to relocalise to the nucleolus in cells depleted of Drosha, however, inhibition of transcription with Actinomycin D inhibited the relocalisation, suggesting DGCR8 is binding nascent RNAs, which are being transported to the nucleolus (Shiohama et al., 2007). These data suggest DGCR8 does indeed bind other RNA substrates when not in a complex with Drosha.

7.1.3 The biological requirement for DGCR8/RRP6 snoRNA regulation

The identification of DGCR8 acting outside of miRNA biogenesis, as a dsRNA adaptor protein, to specifically target mature snoRNAs to the exosome, is novel and of immediate interest. Furthermore, *in vivo* evidence of DGCR8 binding to snoRNAs is interesting as both primary miRNAs and snoRNAs have similar hairpin structures and can be bound by a similar set of proteins (Scott et al., 2009). However, the biological function of the regulation of mature snoRNAs is still unclear. We predict two scenarios which are explained below.

7.1.3.1 (i) DGCR8/RRP6 degrade mature snoRNA at the end of their functional life

To date, no adaptor complexes or ribonucleases have been identified to be solely responsible for the turnover of snoRNAs at the end of their life cycle. The human NEXT complex, composed of MTR4, RBM7 and ZCCHC8 is involved in pre-snoRNA processing. Within our experiments, depletion of RBM7 decreased mature snoRNA levels, confirming the NEXT complex role within pre-snoRNA processing. The other complex known to target snoRNAs for degradation is the TRAMP complex. Within yeast the TRAMP complex, composed of Trf4p/5p Air1p/2p and Mtr4p is responsible for turnover of misfolded snoRNAs, but no evidence has been generated regarding their final degradation. Recently, a homologous human TRAMP complex was identified; comprised of hTRF4-2(Trf4p), ZCCHC7 (Air2p) and MTR4 (Mtr4p), with nucleolar localisation (Lubas et al., 2011). However, to date, there is no strong evidence that has demonstrated hTRAMP participating in snoRNA turnover or other cellular functions. In our experiments, we found the abundance of mature snoRNAs was reduced by ZCCHC7 knockdown,

whereas knockdown of DGCR8 and RRP6 caused an increase. This suggests that in humans, and for the snoRNAs tested, the hTRAMP complex is not involved in turnover of mature snoRNAs but in snoRNA maturation (Fig. 24 B), which has also been suggested in a previous study (Berndt et al., 2012). In contrast, the DGCR8/RRP6-mediated degradation is specific to mature snoRNAs. One explanation for this would be DGCR8 is required for targeting snoRNAs for degradation at the end of their biological life. The terminal degradation of functional snoRNAs has not been investigated thus far, however, like every RNA species within the cell, their turnover must also be regulated.

7.1.3.2 (ii) DGCR8/RRP6 provide a quality control mechanism for misfolded snoRNAs

Quality control mechanisms must sense misfolded snoRNAs and target them to the exosome for degradation. Indeed in yeast, the TRAMP complex has been shown to specifically recognise misfolded tRNAs and target them for exosome-mediated degradation (Vanácová et al., 2005), however the mechanism sensing misfolded snoRNAs remains to be elucidated within all species.

SnoRNA biogenesis is required to liberate snoRNAs embedded within intronic sequences of their host pre-mRNAs. Following this, debranching of the lariat and snoRNP assembly assists snoRNA folding. SnoRNP assembly and binding to the mature snoRNA is a coordinated processes that requires a transition of a non-snoRNA associated, protein only pre-snoRNP complex; to a snoRNA-bound mature snoRNP (Bizarro et al., 2014; McKeegan et al., 2007), which is required for correct maturation. Interestingly, we reported a subtle enrichment of DGCR8 with unspliced snoRNA host genes. Association of DGCR8 to pre-snoRNAs would suggest that these were always destined for degradation. Moreover, the DGCR8 CLIP data revealed that binding of DGCR8 in both H/ACA and C/D box snoRNAs occurs around the basal stem structures that are ~5nt in length (Macias et al., 2012), which is short compared to DGCR8 canonical substrate, pre-miRNAs. The basal stems and conserved box regions of snoRNAs are usually occupied by snoRNP complexes, which are essential for proper folding, stability and activity; therefore it may be counterintuitive to think, due to steric hindrance, that both DGCR8 and snoRNP

complexes could occupy the same location. Therefore, one possible explanation for DGCR8 binding to snoRNAs, could be misassembled snoRNP leading to a misfolded snoRNA. The absence of snoRNP components or changes in secondary structure could enable DGCR8 to bind, sensing those erroneous snoRNAs, and inducing their degradation by recruitment of RRP6. It will therefore be interesting to study if the depletion of pre-snoRNP and mature-snoRNP components enhances the DGCR8-snoRNA interactions, and hence the DGCR8/RRP6 mediated snoRNA degradation.

7.1.4 The biological consequence of increased abundance of snoRNAs

A key question that remains to be addressed is: what is the biological consequence of overexpressing snoRNAs? As the primary function of snoRNAs is to post transcriptionally mature rRNAs, excessive snoRNAs may affect rRNA biogenesis; however, no defects in rRNA processing are reported in *Dgcr8*-/- mESC (Wang et al., 2007). In addition there are no changes in the methylation profile of mature rRNAs (personal comms. Sara Macias). However, concomitant overexpression of Fibrillarin and snoRNAs has been shown to increase tumorigenicity in breast cancer (Su et al., 2014) . In addition, specific snoRNAs are reported to be up regulated in other cancers (Appaiah et al., 2011; Liao et al., 2010; Mei et al., 2012). Taken together, this suggests that individual snoRNAs have proto-oncogenic functions (reviewed in Dupuis-Sandoval et al., 2015).

Recent reports have expanded the functions of snoRNAs beyond rRNA metabolism, and snRNA and tRNA modifications. Bioinformatic analyses of snoRNA guide sequences identified that a subgroup had no predicted complementary RNA targets; these are known as orphan snoRNAs. These orphan snoRNAs are expressed and conserved, which suggests there are functions of snoRNAs yet to be elucidated. For instance, SNORD116 was predicted to have 23 targets within mRNAs (Bazeley et al., 2008; Falaleeva et al., 2015) and hTR/TERC (SCARNA19) functions as an antisense template to initiate telomere elongation as a component of the telomerase RNP (Mitchell et al., 1999). Additionally, SNORD115 has been shown to be processed into sdRNAs (snoRNA derived RNAs) that directly regulate alternative splicing of the serotonin

receptor (Kishore and Stamm, 2006; Kishore et al., 2010). Other sdRNAs from snoRNAs have been shown to harbour miRNA-like functions; highlighting a large repertoire of conserved and abundant sdRNAs which have the potential to function in RNAi (Ender et al., 2008; Scott et al., 2012; Taft et al., 2009). However, another study by Kishore et al. suggested sdRNAs with miRNA-like functions are not as widespread as initially thought (Kishore et al., 2013). In sum, there is a credible amount of evidence suggesting sdRNAs are conserved, however the biological relevance of the majority of these RNAs remains unclear.

An emerging field of research is RNA modifications or epitranscriptomics. Recently, genome wide analysis of the pseudouridylation (Ψ) profile identified that the Ψ modification was also present in mRNAs, and 157 snoRNAs were computationally identified to target mRNAs (Carlile et al., 2014). Interestingly, pseudouridylation was found to increase modified transcripts half-lives. Therefore, it will be interesting to identify what is the effect of overexpressing snoRNAs on the modification status and stability of these mRNAs.

7.1.5 DGCR8 deficiencies and neurological defects

DGCR8 is one of the approximately 30 genes deleted in the DiGeorge syndrome which is caused by a 1.5-3.0Mb hemizygous deletion of chromosome 22q11.2. DiGeorge syndrome is the most common microdeletion syndrome affecting 1 in 4000 births and the clinical manifestations of this congenital disease are cardiac abnormalities, cleft palate, hypocalcaemia/hypoparathyroidism, craniofacial defects and psychiatric illnesses such as schizophrenia. A mutant mouse strain carrying hemizygous 1.5Mb deletion of a region syntenic to the human chromosomal 22q11.2 was engineered in order to gain insight into this congenital disease. This *Dgcr8*+/− mutant mice [*Df(16)A*+/−] displayed a downregulation of miRNA levels by 20% to 70% within the brain. This mouse model also exhibited behavioural, cognitive, and cardiac defects that recapitulated the human condition (Fénelon et al., 2011; Stark et al., 2008). Moreover, the *Df(16)A*+/− mice had abnormalities in the formation of neuronal dendrites, although this was later

attributed to the direct loss of *mir-185*, which also resides within 22q11.2 locus (Xu et al., 2013). Further studies have reported an abundance of neurological problems in *Dgcr8* deficient mice, such as synaptic plasticity and morphology defects (Ouchi et al., 2013; Schofield et al., 2011). Recently, these studies were expanded to flies, which also exhibited neuronal morphology defects (Luhur et al., 2014). Importantly, *pasha/dgcr8* deficiency caused more severe defects than *drosha*. This suggests that the miRNA pathway is not the only pathway affected and suggests that *pasha/dgcr8* may have functions independent of *drosha* that are also important for neuronal morphogenesis. Taken together, these results indicate miRNAs are important for neuronal development and also play a key role in the cognitive defects displayed by DiGeorge syndrome. However, identifying the contribution of the non-canonical functions of DGCR8 (both dependent and independent of Drosha) towards the phenotype seen in *Df(16)A+/−* and DiGeorge syndrome patients would be of extreme interest.

7.2 FUTURE WORK

7.2.1 Genome-wide identification of RRP6 processing hotspots

We performed iCLIP in parallel using WT and CAT version of RRP6. Both iCLIP experiments demonstrated similar sets of RNAs were bound, however, RRP6 CAT binding was enhanced in a subsection of pre-RNAs (5.8S+30nt pre-rRNA and pre-snoRNAs). As this binding is not as enhanced in RRP6 WT iCLIP, this indicates that RRP6 CAT is acting as a dominant negative and is stabilising transcripts usually rapidly degraded by the RRP6. Pre-rRNAs are orders of magnitude less abundant than their respective mature forms, therefore these signals detected within the RRP6-CAT represent highly enriched regions that require the catalytic activity of RRP6 for their removal. By comparing RRP6 WT and CAT profiles in parallel it is possible to generate a genome wide map of RRP6 enriched processing/degradation events. For example, we calculated enrichment of CAT density when compared to WT density over 3' extended snoRNAs (Fig. 40 F). Targets that require RRP6 processing could then be validated using the system outlined below.

7.2.2 *In vivo* validation of RRP6 iCLIP targets

In order to validate the observations described within this chapter and chapter 6, future experiments will address if the RNAs identified by RRP6 CLIP are indeed bona fide substrates. For this purpose, a conventional approach would be to perturb expression of the suspected ribonuclease to identify the stabilised processing intermediates. However, depletion of the exosome associated ribonucleases is challenging, with the exception of RRP6 (van Dijk et al., 2007), and, only the depletion of RRP40, a member of the core exosome, has been shown to effectively inactivate the exosome complex (Lubas et al., 2011). In addition, the complexity of many decay pathways within the cell, mainly represented by the presence of several 3'-5' ribonucleases, suggest a large amount of functional redundancy between them (Copela et al., 2008). Therefore, future work will be

geared towards delineating the processing/decay dependencies of the different RNA species and the mechanisms that stimulate these processes.

During the final stages of my Ph.D., the crystal structure of Rrp6p-Rrp47p-Mtr4p was solved (Schuch et al., 2014). These analyses identified Mtr4p interacts with the exosome through a composite binding site formed by Rrp6p-Rrp47p. As these interactions are also conserved within humans (Schilders et al., 2007), and MTR4 constitutes every nuclear adapter identified thus far; depletion of RRP6 will affect adaptors interacting with the nuclear exosome *in vivo*. Therefore conventional RNAi depletion of RRP6 will not exclusively accumulate intermediates RRP6-dependent processing intermediates, but may also allosterically affect DIS3 substrates.

With this mind, we will employ an alternative approach to study the functional effects of inactivating RRP6 in humans. We will generate isogenic cell lines that express a catalytically inactive and siRNA resistant version of RRP6. Indeed, we have already observed within our iCLIP experiments that RRP6 CAT (D313N; catalytic mutant) stabilised protein:RNA interactions, thus protecting processing/degradation intermediates from degradation by alternative, redundant ribonucleases within the cell. Furthermore, the catalytically inactive RRP6 contains a single point mutation within the exonuclease domain which abrogates it catalytic activity but does not affect its structural functions. Hence, we can circumvent allosteric effects of depleting RRP6.

7.3 CLOSING REMARKS

Phenotypic and transcriptomic profiling experiments designed to identify non-canonical function of the miRNA biogenesis machinery have focused heavily on *dicer* depletion; and used *dgcr8* and *drosha* deficiencies interchangeably to investigate the functions of the Microprocessor complex. To date, there are limited reports detailing phenotypic comparisons of DGCR8, Drosha, Dicer and Ago deficiencies in parallel.

The results presented within this thesis, and other studies, have generated a credible body of evidence identifying that DGCR8 can function independently from Drosha. In addition, the discovery that DGCR8 can function with ribonucleases other than Drosha, gives rise to the interesting hypothesis that DGCR8 may also constitute other complexes that are distinct from the Microprocessor. Taken together, these results highlight an underappreciated role for DGCR8 and underscore the need for further studies profiling *DGCR8, Drosha, Dicer* and *Ago* deficiencies in parallel, in order to delineate further extended functions of the Microprocessor complex components, DGCR8 and Drosha.

Bibliography

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J. *18*, 5399–5410.

Allmang, C., Mitchell, P., Petfalski, E., and Tollervey, D. (2000). Degradation of ribosomal RNA precursors by the exosome. Nucleic Acids Res. *28*, 1684–1691.

Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M., and Eyras, E. (2011). Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. Bioinformatics *27*, 3333–3340.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., et al. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat. Struct. Mol. Biol. *20*, 1367–1376.

Appaiah, H.N., Goswami, C.P., Mina, L.A., Badve, S., Sledge, G.W., Liu, Y., and Nakshatri, H. (2011). Persistent upregulation of U6:SNORD44 small RNA ratio in the serum of breast cancer patients. Breast Cancer Res. *13*, R86.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. Mol. Cell *23*, 841–851.

Artandi, S.E., and DePinho, R.A. (2009). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9–18.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. Cell *152*, 844–858.

Avilion, a a, Piatyszek, M. a, Gupta, J., Shay, J.W., Bacchetti, S., and Greider, C.W. (1996). Human telomerase RNA and telomerase activity in immortal cell lines and tumor tissues. Cancer Res. *56*, 645–650.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. *22*, 2773–2785.

Babiarz, J.E., Hsu, R., Melton, C., Thomas, M., Ullian, E.M., and Blelloch, R. (2011). A role for noncanonical microRNAs in the mammalian brain revealed by phenotypic differences in Dgcr8 versus Dicer1 knockouts and small RNA sequencing. RNA *17*, 1489–1501.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA world of the nucleolus: Two major families of small RNAs defined by different box elements with related functions. Cell *86*, 823–834.

Ballarino, M., Morlando, M., Pagano, F., Fatica, A., and Bozzoni, I. (2005). The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. Mol. Cell. Biol. *25*, 5396–5403.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Battle, D.J., and Doudna, J.A. (2001). The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. RNA *7*, 123–132.

Bazeley, P.S., Shepelev, V., Talebizadeh, Z., Butler, M.G., Fedorova, L., Filatov, V., and Fedorov, A. (2008). snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. Gene *408*, 172–179.

Berndt, H., Harnisch, C., Rammelt, C., Stöhr, N., Zirkel, A., Dohm, J.C., Himmelbauer, H., Tavanez,

J.-P., Hüttelmaier, S., and Wahle, E. (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. RNA *18*, 958–972.

Bernstein, E., Kim, S.Y., Carmell, M. a, Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A. a, Elledge, S.J., Anderson, K. V, and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat. Genet. *35*, 215–217.

Bezman, N. a, Cedars, E., Steiner, D.F., Blelloch, R., Hesslein, D.G.T., and Lanier, L.L. (2010). Distinct requirements of microRNAs in NK cell activation, survival, and function. J. Immunol. *185*, 3835–3846.

Bizarro, J., Charron, C., Boulon, S., Westman, B., Pradet-Balade, B., Vandermoere, F., Chagot, M.-E., Hallais, M., Ahmad, Y., Leonhardt, H., et al. (2014). Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. J. Cell Biol. *207*, 463–480.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. Cell *139*, 547–559.

Broughton, J.P., and Pasquinelli, A.E. (2013). Identifying argonaute binding sites in caenorhabditis elegans using iCLIP. Methods *63*, 119–125.

Callicott, R.J., and Womack, J.E. (2006). Real-time PCR assay for measurement of mouse telomeres. Comp. Med. *56*, 17–22.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W. V (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143–146.

Di Carlo, V., Grossi, E., Laneve, P., Morlando, M., Dini Modigliani, S., Ballarino, M., Bozzoni, I., and Caffarelli, E. (2013). TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation. Mol. Neurobiol. *48*, 952–963.

Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J. *17*, 3726–3737.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584–589.

Cheng, T., Wang, Z., Liao, Q., Zhu, Y., Zhou, W., Xu, W., and Qiu, Z. (2014). Article MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8 / Drosha Complex. Dev. Cell *28*, 547–560.

Chong, M.M.W., Rasmussen, J.P., Rudensky, A.Y., Rundensky, A.Y., and Littman, D.R. (2008). The RNAseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. J. Exp. Med. *205*, 2005–2017.

Chong, M.M.W., Zhang, G., Cheloufi, S., Neubert, T. a, Hannon, G.J., and Littman, D.R. (2010). Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. *24*, 1951–1960.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W.S., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA *15*, 2147–2160.

Copela, L. a, Fernandez, C.F., Sherrer, R.L., and Wolin, S.L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA *14*, 1214–1227.

Costello, J.L., Stead, J. a., Feigenbutz, M., Jones, R.M., and Mitchell, P. (2011). The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation. J. Biol. Chem. *286*, 4535–4543.

Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D., and Bogenhagen, D.F. (1983). Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell *34*, 829–835.

Cristofari, G., and Lingner, J. (2006). Telomere length homeostasis requires that telomerase levels are limiting. EMBO J. *25*, 565–574.

Darzacq, X., Jády, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: A novel class of 2′-O-methylation and pseudouridylation guide RNAs. EMBO J. *21*, 2746–2756.

Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H., and Meier, U.T. (2006). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J. Cell Biol. *173*, 207–218.

Denli, A.M., Tops, B.B.J., Plasterk, R.H. a, Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231–235.

Dez, C., Henras, a, Faucon, B., Lafontaine, D., Caizergues-Ferrer, M., and Henry, Y. (2001). Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. Nucleic Acids Res. *29*, 598–603.

Dhir, A., Dhir, S., Proudfoot, N.J., and Jopling, C.L. (2015). Microprocessor mediates transcriptional termination in genes encoding long noncoding microRNAs. Nat. Struct. Mol. Biol. *22*.

Dieci, G., Preti, M., and Montanini, B. (2009). Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics *94*, 83–88.

van Dijk, E.L., Schilders, G., and Pruijn, G.J.M. (2007). Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. RNA *13*, 1027–1035.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

Dupuis-Sandoval, F., Poirier, M., and Scott, M.S. (2015). The emerging landscape of small nucleolar RNAs in cell biology. Wiley Interdiscip. Rev. RNA *6*, n/a-n/a.

Dziembowski, A., Lorentzen, E., Conti, E., and Séraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat. Struct. Mol. Biol. *14*, 15–22.

Egecioglu, D.E., Henras, A.K., and Chanfreau, G.F. (2006). Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. RNA *12*, 26–32.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Mol. Cell *32*, 519–528.

Falaleeva, M., Surface, J., Shen, M., de la Grange, P., and Stamm, S. (2015). SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. Gene *572*, 266–273.

Falk, S., Weir, J.R., Hentschel, J., Reichelt, P., Bonneau, F., and Conti, E. (2014). The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. Mol. Cell *55*, 856–867.

Faller, M., Matsunaga, M., Yin, S., Loo, J. a, and Guo, F. (2007). Heme is involved in microRNA processing. Nat. Struct. Mol. Biol. *14*, 23–29.

Feigenbutz, M., Jones, R., Besong, T.M.D., Harding, S.E., and Mitchell, P. (2013a). Assembly of the yeast exoribonuclease Rrp6 with its associated cofactor Rrp47 occurs in the nucleus and is critical for the controlled expression of Rrp47. J. Biol. Chem. *288*, 15959–15970.

Feigenbutz, M., Garland, W., Turner, M., and Mitchell, P. (2013b). The exosome cofactor Rrp47 is critical for the stability and normal expression of its associated exoribonuclease Rrp6 in Saccharomyces cerevisiae. PLoS One *8*, e80752.

Fénelon, K., Mukai, J., Xu, B., Hsu, P.-K., Drew, L.J., Karayiorgou, M., Fischbach, G.D., Macdermott, A.B., and Gogos, J. a (2011). Deficiency of Dgcr8, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. *108*, 4447–4452.

Filipowicz, W., and Pogačić, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Curr. Opin. Cell Biol. 319–327.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. *9*, 102–114.

Finnegan, E.F., and Pasquinelli, A.E. (2013). MicroRNA biogenesis: regulating the regulators. Crit. Rev. Biochem. Mol. Biol. *48*, 51–68.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res *42*, D749-55.

Flynt, A.S., Greimann, J.C., Chung, W.-J., Lima, C.D., and Lai, E.C. (2010). MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol. Cell *38*, 900–907.

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and di Fagagna, F. d'Adda (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature *488*, 1–7.

Friedländer, M.R., Lizano, E., Houben, A.J., Bezdan, D., Báñez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B., González, J., Chen, K.C., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. Genome Biol. *15*, R57.

Fukunaga, R., Han, B.W., Hung, J.-H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Dicer Partner Proteins Tune the Length of Mature miRNAs in Flies and Mammals. Cell *151*, 533–546.

Gelpi, C., Algueró, a, Angeles Martinez, M., Vidal, S., Juarez, C., and Rodriguez-Sanchez, J.L. (1990). Identification of protein components reactive with anti-PM/Scl autoantibodies. Clin. Exp. Immunol. *81*, 59–64.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. *10*, 94–108.

Ghorai, A., and Ghosh, U. (2014). miRNA gene counts in chromosomes vary widely in a species and biogenesis of miRNA largely depends on transcription or post-transcriptional processing of coding genes. Front. Genet. *5*, 1–11.

Graves, R. a, and Marzluff, W.F. (1984). Rapid reversible changes in the rate of histone gene transcription and histone mRNA levels in mouse myeloma cells. Mol. Cell. Biol. *4*, 351–357.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R.

(2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. *32*, D109–D111.

Gromak, N., Dienstbier, M., Macias, S., Plass, M., Eyras, E., Cáceres, J.F., and Proudfoot, N.J. (2013). Drosha Regulates Gene Expression Independently of RNA Cleavage Function. Cell Rep. 1499–1510.

Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. Mol. Cell *32*, 247–258.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. *15*, 509–524.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. J. Vis. Exp.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC–ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. Nat. Struct. Mol. Biol. *20*, 1358–1366.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. *18*, 3016–3027.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887–901.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.-K.K., Yeom, K.-H.H., Yang, W.-Y.Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. Cell *136*, 75–84.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M. a (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA 673–695.

Havens, M. a., Reich, A. a., and Hastings, M.L. (2014). Drosha Promotes Splicing of a Pre-microRNA-like Alternative Exon. PLoS Genet. *10*.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell *153*, 654–665.

Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J.L., and Cáceres, J.F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. Nat. Struct. Mol. Biol. *20*, 1173–1181.

Herbert, K.M., Pimienta, G., DeGregorio, S.J., Alexandrov, A., and Steitz, J. a. (2013). Phosphorylation of DGCR8 Increases Its Intracellular Stability and Induces a Progrowth miRNA Profile. Cell Rep. *5*, 1070–1081.

Hoefig, K.P., Rath, N., Heinz, G. a, Wolf, C., Dameris, J., Schepers, A., Kremmer, E., Ansel, K.M., and Heissmeyer, V. (2013). Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. Nat. Struct. Mol. Biol. *20*, 73–81.

van Hoof, A., Lennertz, P., Parker, R., and Parker, R.O.Y. (2000a). Yeast Exosome Mutants Accumulate 3 ′ -Extended Polyadenylated Forms of U4 Small Nuclear RNA and Small Nucleolar RNAs.

van Hoof, A., Lennertz, P., and Parker, R. (2000b). Three conserved members of the RNase D family

have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J. *19*, 1357–1365.

Houseley, J., and Tollervey, D. (2006). Yeast Trf5p is a nuclear poly(A) polymerase. EMBO Rep. *7*, 205–211.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat. Rev. Mol. Cell Biol. *7*, 529–539.

Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., Stefl, R., and Vanacova, S. (2015). RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. Nucleic Acids Res. *43*, 4236–4248.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, G.M., Jarmolowski, A., Strycj, J.C.R., and Fournier, M.J. (1992). Accumulation of U14 Small Nuclear RNA in Saccharomyces cerevisiae Requires Box C , Box D , and a 5 ', 3 ' Terminal Stem. Mol. Cell. Biol. *12*, 4456–4463.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2013). iCLIP: Protein-RNA interactions at nucleotide resolution. Methods.

Jackson, R.N., Klauer, a A., Hintze, B.J., Robinson, H., van Hoof, A., and Johnson, S.J. (2010). The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. EMBO J. *29*, 2205–2216.

Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. J. Cell Biol. *164*, 647–652.

Januszyk, K., and Lima, C.D. (2014). The eukaryotic RNA exosome. Curr. Opin. Struct. Biol. *24C*, 132–140.

Januszyk, K., Liu, Q., and Lima, C.D. (2011). Activities of human RRP6 and structure of the human RRP6 catalytic domain. RNA *17*, 1566–1577.

Kadener, S., Rodriguez, J., Abruzzi, K.C., Khodor, Y.L., Sugino, K., Marr, M.T., Nelson, S., and Rosbash, M. (2009). Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. RNA *15*, 537–545.

Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B.D., Fowler, B.J., Cho, W.G., Kleinman, M.E., Ponicsan, S.L., Hauswirth, W.W., Chiodo, V. a, et al. (2011). DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature *471*, 325–330.

Kanellopoulou, C., Muljo, S. a, Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. *19*, 489–501.

Kataoka, N., Fujita, M., and Ohno, M. (2009). Functional association of the Microprocessor complex with the spliceosome. Mol. Cell. Biol. *29*, 3243–3254.

Kawahara, Y., and Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc. Natl. Acad. Sci. U. S. A. *109*, 3347–3352.

Kerpedjiev, P., Frellsen, J., Lindgreen, S., and Krogh, A. (2014). Adaptable probabilistic mapping of short reads using position specific scoring matrices. BMC Bioinformatics *15*, 100.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate

alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Recpetor 2C. Science *311*, 230–232.

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Hum. Mol. Genet. *19*, 1153–1164.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat. Methods *8*, 559–564.

Kishore, S., Gruber, A.R., Jedlinski, D.J., Syed, A.P., Jorjani, H., and Zavolan, M. (2013). Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. Genome Biol. *14*, R45.

Klauer, a A., and van Hoof, A. (2013). Genetic interactions suggest multiple distinct roles of the arch and core helicase domains of Mtr4 in Rrp6 and exosome function. Nucleic Acids Res. *41*, 533–541.

Knuckles, P., Vogt, M. a, Lugert, S., Milo, M., Chong, M.M.W., Hautbergue, G.M., Wilson, S. a, Littman, D.R., and Taylor, V. (2012). Drosha regulates neurogenesis by controlling Neurogenin 2 expression independent of microRNAs. Nat. Neurosci. *15*, 962–969.

Kobayashi, H., and Tomari, Y. (2015). RISC assembly: Coordination between small RNAs and Argonaute proteins. Biochim. Biophys. Acta - Gene Regul. Mech.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J. Vis. Exp. 1–7.

Koseoglu, M.M., Graves, L.M., and Marzluff, W.F. (2008). Phosphorylation of threonine 61 by cyclin a/Cdk1 triggers degradation of stem-loop binding protein at the end of S phase. Mol. Cell. Biol. *28*, 4469–4479.

Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief. Bioinform. *14*, 144–161.

de la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3'end formation of 5.8 S …. EMBO J. *17*, 1128–1140.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713–724.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The Human DiGeorge Syndrome Critical Region Gene 8 and Its D . melanogaster Homolog Are Required for miRNA Biogenesis. *14*, 2162–2167.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.

Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993–996.

Lee, G., Bratkowski, M. a., Ding, F., Ke, a., and Ha, T. (2012). Elastic Coupling Between RNA Degradation and Unwinding by an Exoribonuclease. Science (80-.). *336*, 1726–1729.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C . elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to & II-14. Cell *75*, 843–854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., and Yim, J. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 1–5.

Lee, Y., Hur, I., Park, S.-Y., Kim, Y.-K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO J. *25*, 522–532.

Liang, S., Hitomi, M., Hu, Y.H., Liu, Y., and Tartakoff, A.M. (1996). A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. Mol. Cell. Biol. *16*, 5139–5146.

Liao, J., Yu, L., Mei, Y., Guarnera, M., Shen, J., Li, R., Liu, Z., and Jiang, F. (2010). Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol. Cancer *9*, 198.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003a). Vertebrate microRNA genes. Science *299*, 1540.

Lim, L.P., Lim, L.P., Lau, N.C., Lau, N.C., Weinstein, E.G., Weinstein, E.G., Abdelhakim, A., Abdelhakim, A., Yekta, S., Yekta, S., et al. (2003b). The microRNAs of Caenorhabditis elegans. Genes Dev. 991–1008.

Lin, Y.-T., and Sullivan, C.S. (2011). Expanding the role of Drosha to the regulation of viral gene expression. Proc. Natl. Acad. Sci. U. S. A. *108*, 11229–11234.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223–1237.

Liu, X., Zheng, Q., Vrettos, N., Maragkakis, M., Alexiou, P., Gregory, B.D., and Mourelatos, Z. (2014). A MicroRNA Precursor Surveillance System in Quality Control of MicroRNA Synthesis. Mol. Cell *3*, 1–12.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the RNase II Nuclease Family. Mol. Cell *29*, 717–728.

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. Mol. Cell *43*, 624–637.

Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H., and Dziembowski, A. (2013). Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. EMBO J. *32*, 1855–1868.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. Cell Rep. *10*, 178–192.

Luhur, a., Chawla, G., Wu, Y.-C., Li, J., and Sokol, N.S. (2014). Drosha-independent DGCR8/Pasha pathway regulates neuronal morphogenesis. Proc. Natl. Acad. Sci. *2013*, 1421–1426.

Lykke-Andersen, S., Brodersen, D.E., and Jensen, T.H. (2009). Origins and activities of the eukaryotic exosome. J. Cell Sci. *122*, 1487–1494.

Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J.F. (2012). DGCR8 HITS-

CLIP reveals novel functions for the Microprocessor. Nat. Struct. Mol. Biol.

MacRae, I.J., and Doudna, J. a (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr. Opin. Struct. Biol. *17*, 138–145.

Makino, D.L., Schuch, B., Stegmann, E., Baumgärtner, M., Basquin, C., and Conti, E. (2015). RNA degradation paths in a 12-subunit nuclear exosome complex. Nature.

Malet, H., Topf, M., Clare, D.K., Ebert, J., Bonneau, F., Basquin, J., Drazkowska, K., Tomecki, R., Dziembowski, A., Conti, E., et al. (2010). RNA channelling by the eukaryotic exosome. EMBO Rep. *11*, 936–942.

Maraia, R.J., and Lamichhane, T.N. (2011). 3' processing of eukaryotic precursor tRNAs.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. *9*, 843–854.

Matera, a G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E., and Watkins, N.J. (2007). A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. Mol. Cell. Biol. *27*, 6782–6793.

Mei, Y.-P., Liao, J.-P., Shen, J., Yu, L., Liu, B.-L., Liu, L., Li, R.-Y., Ji, L., Dorsey, S.G., Jiang, Z.-R., et al. (2012). Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. Oncogene *31*, 2794–2804.

Melamed, Z., Levy, A., Ashwal-Fluss, R., Lev-Maor, G., Mekahel, K., Atias, N., Gilad, S., Sharan, R., Levy, C., Kadener, S., et al. (2013). Alternative Splicing Regulates Biogenesis of miRNAs Located across Exon-Intron Junctions. Mol. Cell *50*, 869–881.

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Mol. Cell. Biol. *28*, 5446–5457.

Mitchell, P., and Tollervey, D. (1996). The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. Genes Dev. 502–513.

Mitchell, J.R., Cheng, J., and Collins, K. (1999). A Box H / ACA Small Nucleolar RNA-Like Domain at the Human Telomerase RNA 3 ′ End. Mol. Cell. Biol. *19*, 567–576.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3' to 5' exoribonucleases. Cell *91*, 457–466.

Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003). Rrp47p Is an Exosome-Associated Protein Required for the 3 J Processing of Stable RNAs. *23*, 6982–6992.

Mori, M., Triboulet, R., Mohseni, M., Schlegelmilch, K., Shrestha, K., Camargo, F.D., and Gregory, R.I. (2014). Hippo signaling regulates microprocessor and links cell-density-dependent mirna biogenesis to cancer. Cell *156*, 893–906.

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nat. Struct. Mol. Biol. *15*, 902–909.

Morlando, M., Dini Modigliani, S., Torrelli, G., Rosa, A., Di Carlo, V., Caffarelli, E., and Bozzoni, I. (2012). FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. EMBO J. *31*, 4502–4510.

Mullen, T.E., and Marzluff, W.F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev. *22*, 50–65.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *102*, 12135–12140.

Nemlich, Y., Greenberg, E., Ortenberg, R., Besser, M.J., Barshack, I., Jacob-hirsch, J., Jacoby, E., Eyal, E., Rivkin, L., Prieto, V.G., et al. (2013). MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. *123*.

Nguyen, T.A.A., Jo, M.H.H., Choi, Y.-G., Park, J., Kwon, S.C.C., Hohng, S., Kim, V.N.N., and Woo, J.-S. (2015). Functional Anatomy of the Human Microprocessor. Cell *161*, 1374–1387.

Ni, J., Tien, A.L., and Fournier, M.J. (1997). Small Nucleolar RNAs Direct Site-Specific Synthesis of Pseudouridine in Ribosomal RNA. Cell *89*, 565–573.

Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y., and Tsukihara, T. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. Science *326*, 1275–1279.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell *130*, 89–100.

Ouchi, Y., Banno, Y., Shimizu, Y., Ando, S., Hasegawa, H., Adachi, K., and Iwamoto, T. (2013). Reduced Adult Hippocampal Neurogenesis and Working Memory Deficits in the Dgcr8-Deficient Mouse Model of 22q11.2 Deletion-Associated Schizophrenia Can Be Rescued by IGF2. J. Neurosci. *33*, 9408–9419.

Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature *475*, 201–205.

Pasquinelli, A.E., McCoy, A., Jiménez, E., Saló, E., Ruvkun, G., Martindale, M.Q., and Baguñà, J. (2003). Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: A role in life history evolution? Evol. Dev. *5*, 372–378.

Petfalski, E., Dandekar, T., Henry, Y., and Tollervey, D. (1998). Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. Mol. Cell. Biol. *18*, 1181–1189.

Pogacić, V., Dragon, F., and Filipowicz, W. (2000). Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. Mol. Cell. Biol. *20*, 9028–9040.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Quick-Cleveland, J., Jacob, J.P., Weitz, S.H., Shoffner, G., Senturia, R., and Guo, F. (2014). The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. Cell Rep. *7*, 1994–2005.

Richard, P., Kiss, A.M., Darzacq, X., Kiss, T., Yang, P.K., Hoareau, C., Froment, C., Monsarrat, B., Henry, Y., and Chanfreau, G. (2005). Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol. Cell. Biol. *25*, 2540–2549.

Richardson, J.E. (2006). fjoin: simple and efficient computation of feature overlaps. J. Comput. Biol. *13*, 1457–1464.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in S. cerevisiae. Mol. Cell *36*, 88–98.

Roth, B.M., Ishimaru, D., and Hennig, M. (2013). The core Microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a non-specific RNA-binding protein. J. Biol. Chem. *8*, 26785–26799.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.

Rybak-wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A Variety of Dicer Substrates in Human and C . elegans. Cell *159*, 1153–1167.

Schilders, G., Raijmakers, R., Raats, J.M.H., and Pruijn, G.J.M. (2005). MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. Nucleic Acids Res. *33*, 6795–6804.

Schilders, G., van Dijk, E., and Pruijn, G.J.M. (2007). C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing. Nucleic Acids Res. *35*, 2564–2572.

Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. Trends Biochem. Sci. *33*, 501–510.

Schneider, C., Anderson, J.T., and Tollervey, D. (2007). The Exosome Subunit Rrp44 Plays a Direct Role in RNA Substrate Recognition. Mol. Cell *27*, 324–331.

Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res. *37*, 1127–1140.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012). Transcriptome-wide analysis of exosome targets. Mol. Cell *48*, 422–433.

Schofield, C.M., Hsu, R., Barker, A.J., Gertz, C.C., Blelloch, R., and Ullian, E.M. (2011). Monoallelic deletion of the microRNA biogenesis gene Dgcr8 produces deficits in the development of excitatory synaptic transmission in the prefrontal cortex. Neural Dev. *6*, 11.

Schuch, B., Feigenbutz, M., Makino, D.L., Falk, S., Basquin, C., Mitchell, P., and Conti, E. (2014). The Exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. EMBO J. *33*, 2737–2881.

Scott, M.S., Avolio, F., Ono, M., Lamond, A.I., and Barton, G.J. (2009). Human miRNA precursors with box H/ACA snoRNA features. PLoS Comput. Biol. *5*.

Scott, M.S., Troshin, P. V, and Barton, G.J. (2011). NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. BMC Bioinformatics *12*, 317.

Scott, M.S., Ono, M., Yamada, K., Endo, A., Barton, G.J., and Lamond, A.I. (2012). Human box C/D snoRNA processing conservation across multiple cell types. Nucleic Acids Res. *40*, 3676–3688.

Seong, Y., Lim, D.-H., Kim, a., Seo, J.H., Lee, Y.S., Song, H., and Kwon, Y.-S. (2014). Global identification of target recognition and cleavage by the Microprocessor in human ES cells. Nucleic Acids Res. *42*, 12806–12821.

Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. Exp. Cell Res. *313*, 4196–4207.

Skowronek, E.W.A., Grzechnik, P., Marchfelder, A., and Kufel, J. (2013). tRNA 3′ processing in yeast involves tRNase Z, Rex1, and Rrp6. 115–130.

Slevin, M.K., Meaux, S., Welch, J.D., Bigler, R., Miliani de Marval, P.L., Su, W., Rhoads, R.E., Prins,

J.F., and Marzluff, W.F. (2014). Deep Sequencing Shows Multiple Oligouridylations Are Required for 3' to 5' Degradation of Histone mRNAs on Polyribosomes. Mol. Cell *53*, 1020–1030.

Sloan, K.E., Schneider, C., and Watkins, N.J. (2012). Comparison of the yeast and human nuclear exosome complexes. Biochem. Soc. Trans. *40*, 850–855.

Sloan, K.E., Mattijssen, S., Lebaron, S., Tollervey, D., Pruijn, G.J.M., and Watkins, N.J. (2013). Both endonucleolytic and exonucleolytic cleavage mediate ITS1 removal during human ribosomal RNA processing. J. Cell Biol. *200*, 577–588.

Stark, K.L., Xu, B., Bagchi, A., Lai, W.-S., Liu, H., Hsu, R., Wan, X., Pavlidis, P., Mills, A. a, Karayiorgou, M., et al. (2008). Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat. Genet. *40*, 751–760.

Stead, J. a., Costello, J.L., Livingstone, M.J., and Mitchell, P. (2007). The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. Nucleic Acids Res. *35*, 5556–5567.

Su, H., Xu, T., Ganapathy, S., Shadfan, M., Long, M., Huang, T.H.-M., Thompson, I., and Yuan, Z.-M. (2014). Elevated snoRNA biogenesis is essential in breast cancer. Oncogene *33*, 1348–1358.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol. *13*, R67.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., Ambrogio, A.D., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature *519*, 491–494.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. Curr. Biol. *20*, 271–277.

Symmons, M.F., Jones, G.H., and Luisi, B.F. (2000). A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. Structure *8*, 1215–1226.

Taft, R.J., Glazov, E. a, Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. RNA *15*, 1233–1240.

Tang, X., Zhang, Y., Tucker, L., and Ramratnam, B. (2010). Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. Nucleic Acids Res. *38*, 6610–6619.

Tang, X., Li, M., Tucker, L., and Ramratnam, B. (2011). Glycogen synthase kinase 3 beta (GSK3-beta) phosphorylates the RNAase III enzyme Drosha at S300 and S302. PLoS One *6*.

Tang, X., Wen, S., Zheng, D., Tucker, L., Cao, L., Pantazatos, D., Moss, S.F., and Ramratnam, B. (2013). Acetylation of Drosha on the N-Terminus Inhibits Its Degradation by Ubiquitination. PLoS One *8*.

Teta, M., Choi, Y.S., Okegbe, T., Wong, G., Tam, O.H., Chong, M.M.W., Seykora, J.T., Nagy, A., Littman, D.R., Andl, T., et al. (2012). Inducible deletion of epidermal Dicer and Drosha reveals multiple functions for miRNAs in postnatal skin. Development *139*, 1405–1416.

Theimer, C. a., Jády, B.E., Chim, N., Richard, P., Breece, K.E., Kiss, T., and Feigon, J. (2007). Structural and Functional Characterization of Human Telomerase RNA Processing and Cajal Body Localization Signals. Mol. Cell *27*, 869–881.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription

Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. Mol. Cell *23*, 853–864.

Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P., et al. (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J. *29*, 2342–2357.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature *459*, 1010–1014.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. Mol. Cell *55*, 467–481.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376–386.

Urlaub, H., Hartmuth, K., and Lührmann, R. (2002). A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. Methods *26*, 170–181.

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. *3*, e189.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *15*, 795–804.

Wada, T., Kikuchi, J., and Furukawa, Y. (2012). Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. EMBO Rep. *13*, 142–149.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., et al. (2012). Microprocessor, Setx, Xrn2, and Rrp6 Co-operate to Induce Premature Termination of Transcription by RNAPII. Cell *150*, 1147–1157.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat. Genet. *39*, 380–385.

Wasmuth, E. V., and Lima, C.D. (2012). Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel. Mol. Cell *48*, 133–144.

Wasmuth, E. V, Januszyk, K., and Lima, C.D. (2014). Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature *511*, 435–439.

Weir, J.R., Bonneau, F., Hentschel, J., and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. Proc. Natl. Acad. Sci. U. S. A. *107*, 12139–12144.

Weitz, S.H., Gong, M., Barr, I., Weiss, S., and Guo, F. (2014). Processing of microRNA primary transcripts requires heme in mammalian cells. Proc. Natl. Acad. Sci. U. S. A.

Xu, B., Hsu, P.K., Stark, K.L., Karayiorgou, M., and Gogos, J.A. (2013). Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. Cell *152*, 262–275.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F.,

Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yang, J.-S., and Lai, E.C. (2011). Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol. Cell *43*, 892–903.

Yao, C., Sasaki, H.M., Ueda, T., Tomari, Y., and Tadakuma, H. (2015). Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex. Mol. Cell *59*, 125–132.

Yeom, K.-H.H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res. *34*, 4622–4629.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011–3016.

Yi, R., Pasolli, H.A., Landthaler, M., Hafner, M., Ojo, T., Sheridan, R., Sander, C., O'Carroll, D., Stoffel, M., Tuschl, T., et al. (2009). DGCR8-dependent microRNA biogenesis is essential for skin development. Proc. Natl. Acad. Sci. U. S. A. *106*, 498–502.

Yoda, M., Cifuentes, D., Izumi, N., Sakaguchi, Y., Suzuki, T., Giraldez, A.J., and Tomari, Y. (2013). Poly(A)-specific ribonuclease mediates 3'-end trimming of argonaute2-cleaved precursor microrna. Cell Rep. *5*, 715–726.

Zeng, Y., and Cullen, B.R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res. *32*, 4776–4785.

Zhu, C., Chen, C., Huang, J., Zhang, H., Zhao, X., Deng, R., Dou, J., Jin, H., Chen, R., Xu, M., et al. (2015). SUMOylation at K 707 of DGCR8 controls direct function of primary microRNA. Nucleic Acids Res. *43*, 7945–7960.

Appendix

>DGCR8 1-773
METDESPSPLPCGPAGEAVMESRARPFQALPREQSPPPPLQTSSGAEVMDVGSGGDGQSELPAEDPFNFYGASLLSKGSFSKG
RLLIDPNCSGHSPRTARHAPAVRKFSPDLKLLKDVKISVSFTESCRSKDRKVLYTGAERDVRAECGLLLSPVSGDVHACPFGG
SVGDGVGIGGESADKKDEENELDQEKRVEYAVLDELEDFTDNLELDEEGAGGFTAKAIVQRDRVDEEALNFPYEDDFDNDVDA
LLEEGLCAPKKRRTEEKYGGDSDHPSDGETSVQPMMTKIKTVLKSRGRPPTEPLPDGWIMTFHNSGVPVYLHRESRVVTWSRP
YFLGTGSIRKHDPPLSSIPCLHYKKMKDNEEREQSSDLTPSGDVSPVKPLSRSAELEFPLDEPDSMGADPGPPDEKDPLGAEA
APGALGQVKAKVEVCKDESVDLEEFRSYLEKRFDFEQVTVKKFRTWAERRQFNREMKRKQAESERPILPANQKLITLSVQDAP
TKKEFVINPNGKSEVCILHEYMQRVLKVRPVYNFFECENPSEPFGASVTIDGVTYGSGTASSKKLAKNKAARATLEILIPDFV
KQTSEEKPKDSEELEYFNHISIEDSRVYELTSKAGLLSPYQILHECLKRNHGMGDTSIKFEVVPGKNQKSEYVMACGKHTVRG
WCKNKRVGKQLASQKILQLLHPHVKNWGSLLRMYGRESSKMVKQETSDKSVIELQQYAKKNKPNLHILSKLQEEMKRLAEERE
ETRKKPKMSIVASAQPGGEPLCTVDV

Appendix 1 |DGCR8 Protein Sequence. Green = Rhed Domain; Blue = WW domain; Red = dsRBD1 & 2; white text = Nucleolar localisation Signal (NoLs)

Appendix 2 | Epitope tagged DGCR8 and Drosha IP-MS data table (Next Page). Detailed within this table are the individual results from the IP-MS experiments. Enrichment was calculated as Log2(IP Mascot Score/IP Control Mascot), and a value >1 represents a specific enrichment over the appropriate control IP (Green shaded boxes in Enrichment column), whereas a value ≤1 represents non-specific enrichment over the control IP (shaded red boxes in enrichment column). From these values, proteins identified in the IP-MS analyses were segregated based on their association to DGCR8 and Drosha (see Group column). Within the "Gene Names" column selected boxes are shaded based on their association with this project and these are as follows: green shading = bait proteins; red shading = Exosome or Exosome-associated proteins; yellow shading = proteins which have been reported to associate with the Microprocessor complex; blue shading = components of the snoRNP. All ribosomal proteins enriched with DGCR8 or Drosha IPs were disqualified from these analyses and can be found at the end of the table in the group "Ribosomal proteins." Abbreviations used within table are as follows: T7= T7 empty vector, T7-D8= T7-DGCR8; F = FLAG, F-D8 = FLAG-DGCR8, F-DR = FLAG-Drosha; Uni. Pep. = Unique Peptides, Cover. = Coverage

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
Q8WYQ5	DGCR8	Microprocessor complex subunit DGCR8	773	85.99	5.95	1.00	0.00	0.00	7451.87	80.08	67.00	59.55	2.20	2.00	3403.04	79.82	59.00	128.36	13.20	9.00	12.86	5.84	1.11	ALL
Q9NRR4	DROSHA	Ribonuclease 3	1374	159.21	7.87	1.00	0.00	0.00	372.14	12.88	13.00	1.00	0.00	0.00	229.45	7.57	9.00	1892.83	40.03	45.00	8.54	7.84	10.89	ALL
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	579	63.67	8.87	1.00	0.00	0.00	178.03	10.19	3.00	1.00	0.00	0.00	107.47	4.66	2.00	143.06	8.64	2.00	7.48	6.75	7.16	ALL
P19338	NCL	Nucleolin	710	76.57	4.70	519.49	21.27	13.00	2055.16	32.54	24.00	322.83	20.99	13.00	819.08	30.99	23.00	684.54	24.08	15.00	1.98	1.34	1.08	ALL
Q8NC51	SERBP1	Plasminogen activator inhibitor 1 RNA-binding protein	408	44.94	8.65	1.00	0.00	0.00	134.00	10.54	3.00	1.00	0.00	0.00	136.43	10.54	3.00	69.92	10.54	3.00	7.07	7.09	6.13	ALL
P09012	SNRPA	zinc finger protein RFP	282	31.26	9.83	1.00	0.00	0.00	76.53	13.12	3.00	1.00	0.00	0.00	44.39	6.38	2.00	56.08	6.38	2.00	6.26	5.47	5.81	ALL
Q96SB4	SRPK1	SRSF protein kinase 1	655	74.28	6.16	1.00	0.00	0.00	234.12	14.50	7.00	1.00	0.00	0.00	88.94	4.43	2.00	89.42	6.56	3.00	7.87	6.47	6.48	ALL
Q13247	SRSF6	Serine/arginine-rich splicing factor 6	344	39.56	11.43	86.08	8.72	2.00	180.89	19.48	4.00	1.00	0.00	0.00	60.40	6.10	2.00	64.26	6.10	2.00	1.07	5.92	6.01	ALL
P67809	YBX1	Nuclease-sensitive element-binding protein 1	324	35.90	9.88	347.17	49.69	9.00	702.37	44.14	7.00	121.33	21.60	3.00	562.99	50.31	9.00	332.76	47.22	7.00	1.02	2.21	1.46	ALL
P16989	YBX3	Y-box-binding protein 3	372	40.07	9.77	1.00	0.00	0.00	285.08	23.12	2.00	1.00	0.00	0.00	193.18	23.12	2.00	143.87	20.70	2.00	8.16	7.59	7.17	ALL
Q9BZE4	GTPBP4	Nucleolar GTP-binding protein 1	634	73.92	9.50	1.00	0.00	0.00	744.65	37.85	22.00	1.00	0.00	0.00	391.10	21.14	12.00	1.00	0.00	0.00	9.54	8.61	0.00	DGCR8 Exclusive
Q9NX58	LYAR	Cell growth-regulating nucleolar protein	379	43.59	9.54	1.00	0.00	0.00	622.06	42.22	18.00	1.00	0.00	0.00	94.86	7.12	2.00	1.00	0.00	0.00	9.28	6.57	0.00	DGCR8 Exclusive
Q7L2E3	DHX30	Putative ATP-dependent RNA helicase DHX30	1194	133.85	8.78	1.00	0.00	0.00	453.14	15.83	16.00	1.00	0.00	0.00	44.87	1.84	2.00	1.00	0.00	0.00	8.82	5.49	0.00	DGCR8 Exclusive
P06899	HIST1H2BJ	Histone H2B type 1-J	126	13.90	10.32	1.00	0.00	0.00	431.08	35.71	4.00	1.00	0.00	0.00	243.32	35.71	2.00	1.00	0.00	0.00	8.75	7.93	0.00	DGCR8 Exclusive
Q7Z2T5	TRMT1L	TRMT1-like protein	733	81.70	7.88	1.00	0.00	0.00	371.87	16.23	8.00	1.00	0.00	0.00	194.04	9.69	4.00	1.00	0.00	0.00	8.54	7.60	0.00	DGCR8 Exclusive
Q9NQT5	EXOSC3	Exosome complex component RRP40	275	29.55	8.10	1.00	0.00	0.00	342.03	49.45	8.00	1.00	0.00	0.00	81.95	26.91	4.00	1.00	0.00	0.00	8.42	6.36	0.00	DGCR8 Exclusive
Q9BYG3	NIFK	MKI67 FHA domain-interacting nucleolar phosphoprotein	293	34.20	9.88	1.00	0.00	0.00	328.24	37.88	7.00	1.00	0.00	0.00	52.27	7.85	2.00	1.00	0.00	0.00	8.36	5.71	0.00	DGCR8 Exclusive
P56537	EIF6	Eukaryotic translation initiation factor 6	245	26.58	4.68	1.00	0.00	0.00	328.07	56.73	7.00	1.00	0.00	0.00	85.47	19.59	3.00	1.00	0.00	0.00	8.36	6.42	0.00	DGCR8 Exclusive
Q9NW13	RBM28	RNA-binding protein 28	759	85.68	9.22	1.00	0.00	0.00	326.24	20.82	14.00	1.00	0.00	0.00	72.44	5.27	4.00	1.00	0.00	0.00	8.35	6.18	0.00	DGCR8 Exclusive
Q06265	EXOSC9	Exosome complex component RRP45	439	48.92	5.29	1.00	0.00	0.00	314.29	23.01	7.00	1.00	0.00	0.00	104.63	5.24	2.00	1.00	0.00	0.00	8.30	6.71	0.00	DGCR8 Exclusive
P42696	RBM34	RNA-binding protein 34	430	48.54	10.11	1.00	0.00	0.00	289.69	24.65	10.00	1.00	0.00	0.00	57.07	5.35	2.00	1.00	0.00	0.00	8.18	5.83	0.00	DGCR8 Exclusive
Q9NVU7	SDAD1	Protein SDA1 homolog	687	79.82	9.25	1.00	0.00	0.00	254.43	12.23	7.00	1.00	0.00	0.00	83.85	5.97	4.00	1.00	0.00	0.00	7.99	6.39	0.00	DGCR8 Exclusive
Q9BU76	MMTAG2	Multiple myeloma tumor-associated protein 2	263	29.39	10.02	1.00	0.00	0.00	251.49	23.95	5.00	1.00	0.00	0.00	86.73	9.89	2.00	1.00	0.00	0.00	7.97	6.44	0.00	DGCR8 Exclusive
Q9BQ67	GRWD1	Glutamate-rich WD repeat-containing protein 1	446	49.39	4.92	1.00	0.00	0.00	238.32	22.87	5.00	1.00	0.00	0.00	160.63	12.33	4.00	1.00	0.00	0.00	7.90	7.33	0.00	DGCR8 Exclusive
Q00577	PURA	Transcriptional activator protein Pur-alpha	322	34.89	6.44	1.00	0.00	0.00	233.41	28.88	5.00	1.00	0.00	0.00	39.69	9.01	2.00	1.00	0.00	0.00	7.87	5.31	0.00	DGCR8 Exclusive
Q9NPD3	EXOSC4	Exosome complex component RRP41	245	26.37	6.52	1.00	0.00	0.00	221.29	21.63	4.00	1.00	0.00	0.00	117.19	17.55	3.00	1.00	0.00	0.00	7.79	6.87	0.00	DGCR8 Exclusive
Q5RKV6	EXOSC6	Exosome complex component MTR3	272	28.22	6.28	1.00	0.00	0.00	189.09	27.94	5.00	1.00	0.00	0.00	91.09	26.47	5.00	1.00	0.00	0.00	7.56	6.51	0.00	DGCR8 Exclusive
Q9NQ29	LUC7L	Putative RNA-binding protein Luc7-like 1	371	43.70	9.92	1.00	0.00	0.00	187.17	15.09	2.00	1.00	0.00	0.00	54.26	4.85	2.00	1.00	0.00	0.00	7.55	5.76	0.00	DGCR8 Exclusive
Q92522	H1FX	Histone H1x	213	22.47	10.76	1.00	0.00	0.00	176.13	25.82	4.00	1.00	0.00	0.00	158.13	17.84	4.00	1.00	0.00	0.00	7.46	7.30	0.00	DGCR8 Exclusive
Q13868	EXOSC2	Exosome complex component RRP4	293	32.77	7.50	1.00	0.00	0.00	162.80	19.11	4.00	1.00	0.00	0.00	95.85	12.97	3.00	1.00	0.00	0.00	7.35	6.58	0.00	DGCR8 Exclusive
Q9BRJ6	C7orf50	Uncharacterized protein C7orf50	194	22.07	9.64	1.00	0.00	0.00	151.57	48.97	6.00	1.00	0.00	0.00	88.88	24.74	3.00	1.00	0.00	0.00	7.24	6.47	0.00	DGCR8 Exclusive
Q9BZI7	UPF3B	Regulator of nonsense transcripts 3B	483	57.73	9.48	1.00	0.00	0.00	133.97	10.77	4.00	1.00	0.00	0.00	58.24	4.76	2.00	1.00	0.00	0.00	7.07	5.86	0.00	DGCR8 Exclusive
Q9UNQ2	DIMT1	Probable dimethyladenosine transferase	313	35.21	9.99	1.00	0.00	0.00	115.42	11.50	3.00	1.00	0.00	0.00	110.16	9.58	3.00	1.00	0.00	0.00	6.85	6.78	0.00	DGCR8 Exclusive
P16403	HIST1H1C	Histone H1.2	213	21.35	10.93	1.00	0.00	0.00	113.91	15.49	4.00	1.00	0.00	0.00	171.21	24.88	6.00	1.00	0.00	0.00	6.83	7.42	0.00	DGCR8 Exclusive
P05455	SSB	Lupus La protein	408	46.81	7.12	1.00	0.00	0.00	107.34	6.13	2.00	1.00	0.00	0.00	134.68	9.56	3.00	1.00	0.00	0.00	6.75	7.07	0.00	DGCR8 Exclusive
P37108	SRP14	Signal recognition particle 14 kDa protein	136	14.56	10.04	1.00	0.00	0.00	93.53	26.47	3.00	1.00	0.00	0.00	188.43	27.21	3.00	1.00	0.00	0.00	6.55	7.56	0.00	DGCR8 Exclusive
P49458	SRP9	Signal recognition particle 9 kDa protein	86	10.11	7.97	1.00	0.00	0.00	93.15	25.58	2.00	1.00	0.00	0.00	107.77	25.58	2.00	1.00	0.00	0.00	6.54	6.75	0.00	DGCR8 Exclusive
Q6PCB5	RSBN1L	Round spermatid basic protein 1-like protein	846	94.81	8.78	1.00	0.00	0.00	83.38	4.49	3.00	1.00	0.00	0.00	70.61	3.07	2.00	1.00	0.00	0.00	6.38	6.14	0.00	DGCR8 Exclusive
Q9BQ75	CMSS1	Protein CMSS1	279	31.86	9.19	1.00	0.00	0.00	78.15	8.96	3.00	1.00	0.00	0.00	41.98	5.73	2.00	1.00	0.00	0.00	6.29	5.39	0.00	DGCR8 Exclusive
P41091	EIF2S3	Eukaryotic translation initiation factor 2 subunit 3	472	51.08	8.40	1.00	0.00	0.00	53.94	6.36	2.00	1.00	0.00	0.00	108.01	14.62	4.00	1.00	0.00	0.00	5.75	6.75	0.00	DGCR8 Exclusive
Q8IUF8	MINA	Bifunctional lysine-specific demethylase and histidyl-hydroxylase	465	52.77	6.70	1.00	0.00	0.00	42.83	4.52	2.00	1.00	0.00	0.00	60.53	6.67	3.00	1.00	0.00	0.00	5.42	5.92	0.00	DGCR8 Exclusive
Q6NXT2	H3F3C	Histone H3.3C	135	15.20	11.11	1.00	0.00	0.00	42.71	11.85	2.00	1.00	0.00	0.00	28.66	11.11	2.00	1.00	0.00	0.00	5.42	4.84	0.00	DGCR8 Exclusive
O76094	SRP72	Signal recognition particle subunit SRP72	671	74.56	9.26	1.00	0.00	0.00	41.61	4.32	2.00	1.00	0.00	0.00	101.14	4.02	2.00	1.00	0.00	0.00	5.38	6.66	0.00	DGCR8 Exclusive
O76021	RSL1D1	Ribosomal L1 domain-containing protein 1	490	54.94	10.13	135.37	8.16	3.00	867.88	38.98	22.00	1.00	0.00	0.00	112.30	5.10	2.00	1.00	0.00	0.00	2.68	6.81	0.00	DGCR8 Exclusive
P56182	RRP1	Ribosomal RNA processing protein 1 homolog A	461	52.81	9.33	71.13	5.42	2.00	218.31	16.49	6.00	1.00	0.00	0.00	56.59	4.56	2.00	1.00	0.00	0.00	1.62	5.82	0.00	DGCR8 Exclusive
P46087	NOP2	Putative ribosomal RNA methyltransferase NOP2	812	89.25	9.23	231.68	7.39	4.00	535.67	21.43	14.00	1.00	0.00	0.00	129.80	6.65	4.00	1.00	0.00	0.00	1.21	7.02	0.00	DGCR8 Exclusive
Q9BQG0	MYBBP1A	Myb-binding protein 1A	1328	148.76	9.28	286.24	8.36	9.00	1862.25	43.07	46.00	1.00	0.00	0.00	232.30	9.94	11.00	1.00	0.00	0.00	2.70	7.86	0.00	DGCR8 Exclusive
Q8TDN6	BRIX1	Ribosome biogenesis protein BRX1 homolog	353	41.37	9.92	65.66	5.10	2.00	421.00	36.26	11.00	1.00	0.00	0.00	91.64	13.88	4.00	1.00	0.00	0.00	2.68	6.52	0.00	DGCR8 Exclusive
O95793	STAU1	Double-stranded RNA-binding protein Staufen homolog 1	577	63.14	9.44	65.60	4.16	2.00	378.00	16.98	8.00	1.00	0.00	0.00	113.03	10.57	5.00	1.00	0.00	0.00	2.53	6.82	0.00	DGCR8 Exclusive
Q9BVP2	GNL3	Guanine nucleotide-binding protein-like 3	549	61.95	9.16	120.23	9.11	3.00	663.17	33.52	14.00	1.00	0.00	0.00	254.30	13.30	7.00	1.00	0.00	0.00	2.46	7.99	0.00	DGCR8 Exclusive
Q9H7B2	RPF2	Ribosome production factor 2 homolog	306	35.56	9.99	39.36	5.23	2.00	210.28	31.37	9.00	1.00	0.00	0.00	43.61	7.84	2.00	1.00	0.00	0.00	2.42	5.45	0.00	DGCR8 Exclusive
Q99848	EBNA1BP2	Probable rRNA-processing protein EBP2	306	34.83	10.10	58.71	8.82	2.00	280.63	37.91	10.00	1.00	0.00	0.00	79.32	11.76	3.00	1.00	0.00	0.00	2.26	6.31	0.00	DGCR8 Exclusive
Q9NR30	DDX21	Nucleolar RNA helicase 2	783	87.29	9.28	278.42	15.58	9.00	1223.21	45.21	30.00	1.00	0.00	0.00	146.10	13.15	8.00	1.00	0.00	0.00	2.14	7.19	0.00	DGCR8 Exclusive
Q01780	EXOSC10	Exosome component 10 RRP6	885	100.77	8.46	94.00	2.71	2.00	387.63	20.56	14.00	1.00	0.00	0.00	118.30	4.63	3.00	1.00	0.00	0.00	2.04	6.89	0.00	DGCR8 Exclusive
Q99459	CDC5L	Cell division cycle 5-like protein	802	92.19	8.18	166.08	7.11	5.00	552.28	26.93	17.00	1.00	0.00	0.00	152.60	7.86	5.00	1.00	0.00	0.00	1.73	7.25	0.00	DGCR8 Exclusive
P06748	NPM1	Nucleophosmin	294	32.55	4.78	286.87	36.39	7.00	788.36	50.00	13.00	150.34	15.31	5.00	408.31	44.90	10.00	270.49	23.81	6.00	1.46	1.44	0.85	DGCR8 Exclusive
Q9NZI8	IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1	577	63.44	9.20	176.24	12.13	5.00	464.78	31.37	13.00	146.57	10.57	4.00	308.91	23.22	11.00	183.46	15.25	6.00	1.40	1.08	0.32	DGCR8 Exclusive
Q9H0D6	XRN2	5'-3' exoribonuclease 2	950	108.51	7.47	74.06	2.84	2.00	155.05	6.00	4.00	1.00	0.00	0.00	145.30	5.47	4.00	1.00	0.00	0.00	1.07	7.18	0.00	DGCR8 Exclusive
Q86V81	ALYREF	THO complex subunit 4	257	26.87	11.15	210.08	22.18	3.00	430.87	36.19	7.00	1.00	0.00	0.00	90.72	11.28	2.00	1.00	0.00	0.00	1.04	6.50	0.00	DGCR8 Exclusive
Q12906	ILF3	Interleukin enhancer-binding factor 3	894	95.28	8.76	317.78	10.07	8.00	651.41	29.08	17.00	83.12	3.24	3.00	418.75	21.36	14.00	77.16	4.47	4.00	1.04	2.33	-0.11	DGCR8 Exclusive
Q8TCU4	ALMS1	Alstrom syndrome protein 1	4167	460.68	6.28	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	28.57	0.38	1.00	0.00	0.00	4.84	FLAG Drosha Only
Q9Y224	C14orf166	UPF0568 protein C14orf166	244	28.05	6.65	65.17	11.48	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	36.45	10.66	1.00	-6.03	0.00	5.19	FLAG Drosha Only
P68400	CSNK2A1	Casein kinase II subunit alpha	391	45.11	7.74	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	110.05	18.16	5.00	0.00	0.00	6.78	FLAG Drosha Only
O15523	DDX3Y	ATP-dependent RNA helicase DDX3Y	660	73.11	7.55	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	142.52	6.67	3.00	0.00	0.00	7.15	FLAG Drosha Only
P15924	DSP	Desmoplakin	2871	331.57	6.81	159.90	1.95	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	36.28	0.77	2.00	-7.32	0.00	5.18	FLAG Drosha Only

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
P55884	EIF3B	Eukaryotic translation initiation factor 3 subunit B	814	92.42	5.00	276.39	21.13	11.00	213.83	9.21	5.00	1.00	0.00	0.00	1.00	0.00	0.00	57.15	3.32	2.00	-0.37	0.00	5.84	FLAG Drosha Only
Q9Y262	EIF3L	Eukaryotic translation initiation factor 3 subunit L	564	66.68	6.34	222.03	12.41	6.00	171.23	7.62	4.00	1.00	0.00	0.00	1.00	0.00	0.00	55.76	3.19	2.00	-0.37	0.00	5.80	FLAG Drosha Only
Q15717	ELAVL1	ELAV-like protein 1	326	36.07	9.17	104.16	10.12	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	51.53	6.13	2.00	-6.70	0.00	5.69	FLAG Drosha Only
Q5D862	FLG2	Filaggrin-2	2391	247.93	8.31	127.91	3.39	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	65.86	2.59	2.00	-7.00	0.00	6.04	FLAG Drosha Only
P35637	FUS	RNA-binding protein FUS	526	53.39	9.36	238.20	19.20	7.00	88.80	7.79	3.00	1.00	0.00	0.00	1.00	0.00	0.00	101.24	7.41	4.00	-1.42	0.00	6.66	FLAG Drosha Only
Q9H840	GEMIN7	Gem-associated protein 7	131	14.53	7.40	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	59.93	35.11	2.00	0.00	0.00	5.91	FLAG Drosha Only
Q14103	HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	355	38.41	7.81	132.25	14.08	3.00	242.95	19.44	4.00	1.00	0.00	0.00	1.00	0.00	0.00	75.31	6.20	2.00	0.88	0.00	6.23	FLAG Drosha Only
P52597	HNRNPF	Heterogeneous nuclear ribonucleoprotein F	415	45.64	5.58	324.82	25.30	4.00	375.70	31.33	5.00	1.00	0.00	0.00	1.00	0.00	0.00	199.15	22.89	4.00	0.21	0.00	7.64	FLAG Drosha Only
P14866	HNRNPL	Heterogeneous nuclear ribonucleoprotein L	589	64.09	8.22	286.16	20.54	9.00	243.24	14.60	6.00	1.00	0.00	0.00	1.00	0.00	0.00	72.69	4.07	2.00	-0.23	0.00	6.18	FLAG Drosha Only
Q92945	KHSRP	Far upstream element-binding protein 2	711	73.07	7.30	271.65	14.06	6.00	67.55	5.06	2.00	1.00	0.00	0.00	1.00	0.00	0.00	94.08	5.20	3.00	-2.01	0.00	6.56	FLAG Drosha Only
P02533	KRT14	Keratin, type I cytoskeletal 14	472	51.53	5.16	317.96	22.25	6.00	1.00	0.00	0.00	187.83	9.96	2.00	1.00	0.00	0.00	380.31	28.18	5.00	-8.31	-7.55	1.02	FLAG Drosha Only
P08779	KRT16	Keratin, type I cytoskeletal 16	473	51.24	5.05	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	339.05	21.99	3.00	0.00	0.00	8.41	FLAG Drosha Only
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	639	65.39	8.00	1387.05	59.00	23.00	189.32	10.64	4.00	464.04	29.11	12.00	470.36	21.91	8.00	945.30	54.62	20.00	-2.87	0.02	1.03	FLAG Drosha Only
P13647	KRT5	Keratin, type II cytoskeletal 5	590	62.34	7.74	394.71	20.68	5.00	1.00	0.00	0.00	161.95	11.02	2.00	1.00	0.00	0.00	328.54	16.61	4.00	-8.62	-7.34	1.02	FLAG Drosha Only
P02538	KRT6A	Keratin, type II cytoskeletal 6A	564	60.01	8.00	393.75	20.57	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	345.04	18.44	3.00	-8.62	0.00	8.43	FLAG Drosha Only
Q7RTS7	KRT74	Keratin, type II cytoskeletal 74	529	57.83	7.71	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	89.93	6.24	2.00	0.00	0.00	6.49	FLAG Drosha Only
P06309	KV205	Ig kappa chain V-II region GM607 (Fragment)	117	12.66	7.12	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	2742.88	17.95	2.00	0.00	0.00	11.42	FLAG Drosha Only
P06313	KV403	Ig kappa chain V-IV region JI	133	14.62	6.58	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	70.84	17.29	2.00	0.00	0.00	6.15	FLAG Drosha Only
Q9Y383	LUC7L2	Putative RNA-binding protein Luc7-like 2	392	46.49	10.01	239.05	20.41	7.00	234.33	19.64	4.00	1.00	0.00	0.00	1.00	0.00	0.00	120.57	10.20	4.00	-0.03	0.00	6.91	FLAG Drosha Only
Q13310	PABPC4	Polyadenylate-binding protein 4	644	70.74	9.26	265.48	14.75	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	223.54	14.13	5.00	-8.05	0.00	7.80	FLAG Drosha Only
P30153	PPP2R1A	Ser/Thr-protein phosphatase 2A 65 kDa reg. subunit A α iso	589	65.27	5.11	315.36	22.41	9.00	74.93	4.75	2.00	1.00	0.00	0.00	1.00	0.00	0.00	69.82	5.09	2.00	-2.07	0.00	6.13	FLAG Drosha Only
Q06830	PRDX1	Peroxiredoxin-1	199	22.10	8.13	97.53	14.57	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	41.45	9.55	2.00	-6.61	0.00	5.37	FLAG Drosha Only
Q8WWY3	PRPF31	U4/U6 small nuclear ribonucleoprotein Prp31	499	55.42	5.78	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	88.29	6.61	3.00	0.00	0.00	6.46	FLAG Drosha Only
P26599	PTBP1	Polypyrimidine tract-binding protein 1	531	57.19	9.17	571.54	32.20	11.00	464.96	28.06	8.00	1.00	0.00	0.00	1.00	0.00	0.00	144.54	11.11	3.00	-0.30	0.00	7.18	FLAG Drosha Only
P61026	RAB10	Ras-related protein Rab-10	200	22.53	8.38	377.89	52.00	7.00	125.62	22.50	2.00	1.00	0.00	0.00	1.00	0.00	0.00	64.26	11.00	2.00	-1.59	0.00	6.01	FLAG Drosha Only
Q96PK6	RBM14	RNA-binding protein 14	669	69.45	9.67	168.85	11.81	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	190.10	13.60	7.00	-7.40	0.00	7.57	FLAG Drosha Only
Q96I25	RBM17	Splicing factor 45	401	44.93	5.97	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	55.29	4.74	2.00	0.00	0.00	5.79	FLAG Drosha Only
P49756	RBM25	RNA-binding protein 25	843	100.12	6.32	41.42	3.44	2.00	54.41	3.44	2.00	1.00	0.00	0.00	1.00	0.00	0.00	74.60	2.61	2.00	0.39	0.00	6.22	FLAG Drosha Only
Q9BWF3	RBM4	RNA-binding protein 4	364	40.29	7.08	250.37	32.69	8.00	245.38	27.75	8.00	1.00	0.00	0.00	1.00	0.00	0.00	125.25	11.81	4.00	-0.03	0.00	6.97	FLAG Drosha Only
Q15459	SF3A1	Splicing factor 3A subunit 1	793	88.83	5.22	119.59	9.21	5.00	67.19	7.31	3.00	1.00	0.00	0.00	1.00	0.00	0.00	58.41	3.66	2.00	-0.83	0.00	5.87	FLAG Drosha Only
Q15393	SF3B3	Splicing factor 3B subunit 3	1217	135.49	5.26	587.17	20.05	19.00	451.85	11.67	11.00	1.00	0.00	0.00	1.00	0.00	0.00	150.41	4.93	5.00	-0.38	0.00	7.23	FLAG Drosha Only
P23246	SFPQ	Splicing factor, proline- and glutamine-rich	707	76.10	9.44	549.25	28.71	17.00	200.39	11.17	7.00	1.00	0.00	0.00	1.00	0.00	0.00	44.24	3.68	2.00	-1.45	0.00	5.47	FLAG Drosha Only
Q8NI27	THOC2	THO complex subunit 2	1593	182.66	8.44	211.54	3.33	5.00	131.19	2.13	3.00	1.00	0.00	0.00	1.00	0.00	0.00	78.96	1.38	2.00	-0.69	0.00	6.30	FLAG Drosha Only
P14373	TRIM27	Zinc finger protein RFP	513	58.45	6.21	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	62.61	5.85	2.00	0.00	0.00	5.97	FLAG Drosha Only
Q71U36	TUBA1A	Tubulin alpha-1A chain	451	50.10	5.06	1946.53	81.60	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1050.85	63.86	2.00	-10.93	0.00	10.04	FLAG Drosha Only
Q9Y5A9	YTHDF2	YTH domain-containing family protein 2	579	62.30	8.79	112.09	7.60	4.00	90.40	7.08	3.00	1.00	0.00	0.00	1.00	0.00	0.00	63.05	6.22	3.00	-0.31	0.00	5.98	FLAG Drosha Only
P63104	YWHAZ	14-3-3 protein zeta/delta	245	27.73	4.79	233.11	33.88	5.00	171.65	30.61	6.00	1.00	0.00	0.00	1.00	0.00	0.00	91.01	10.61	2.00	-0.44	0.00	6.51	FLAG Drosha Only
Q12905	ILF2	Interleukin enhancer-binding factor 2	390	43.04	5.26	482.27	41.03	9.00	936.45	56.41	15.00	80.31	5.38	2.00	378.87	45.90	11.00	92.31	7.95	3.00	0.96	2.24	0.20	FLAG-DGCR8 Only
Q08211	DHX9	ATP-dependent RNA helicase A	1270	140.87	6.84	843.15	25.91	25.00	1619.82	40.24	43.00	82.01	2.99	4.00	423.56	16.93	17.00	163.79	5.75	7.00	0.94	2.37	1.00	FLAG-DGCR8 Only
Q14498	RBM39	RNA-binding protein 39	530	59.34	10.10	235.80	18.49	8.00	432.72	24.72	10.00	1.00	0.00	0.00	108.94	8.11	3.00	1.00	0.00	0.00	0.88	6.77	0.00	FLAG-DGCR8 Only
P22087	FBL	rRNA 2'-O-methyltransferase fibrillarin	321	33.76	10.18	139.24	24.92	5.00	231.45	39.25	5.00	1.00	0.00	0.00	53.36	9.35	2.00	1.00	0.00	0.00	0.73	5.74	0.00	FLAG-DGCR8 Only
Q9Y5Q9	GTF3C3	General transcription factor 3C polypeptide 3	886	101.21	5.07	108.67	6.32	4.00	171.54	4.51	4.00	1.00	0.00	0.00	81.52	3.84	2.00	1.00	0.00	0.00	0.66	6.35	0.00	FLAG-DGCR8 Only
O43390	HNRNPR	Heterogeneous nuclear ribonucleoprotein R	633	70.90	8.13	244.18	15.17	7.00	323.17	27.01	10.00	1.00	0.00	0.00	71.43	5.06	3.00	1.00	0.00	0.00	0.40	6.16	0.00	FLAG-DGCR8 Only
O94906	PRPF6	Pre-mRNA-processing factor 6	941	106.86	8.25	103.74	1.91	2.00	136.45	3.72	3.00	1.00	0.00	0.00	197.63	7.01	5.00	1.00	0.00	0.00	0.40	7.63	0.00	FLAG-DGCR8 Only
P46060	RANGAP1	Ran GTPase-activating protein 1	587	63.50	4.68	251.53	19.59	8.00	254.90	15.16	6.00	1.00	0.00	0.00	92.04	8.18	3.00	1.00	0.00	0.00	0.02	6.52	0.00	FLAG-DGCR8 Only
Q9UHI6	DDX20	Probable ATP-dependent RNA helicase DDX20	824	92.18	6.95	1.00	0.00	0.00	1.00	0.00	0.00	103.17	10.19	7.00	214.90	12.26	8.00	162.90	15.41	11.00	0.00	1.06	0.66	FLAG-DGCR8 Only
O75122	CLASP2	CLIP-associating protein 2	1294	141.05	8.47	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	74.96	1.62	2.00	1.00	0.00	0.00	0.00	6.23	0.00	FLAG-DGCR8 Only
P01616	KV203	Ig kappa chain V-II region MIL	112	12.05	9.29	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	2832.95	14.29	3.00	1.00	0.00	0.00	0.00	11.47	0.00	FLAG-DGCR8 Only
P05198	EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	315	36.09	5.08	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	79.39	6.35	2.00	1.00	0.00	0.00	0.00	6.31	0.00	FLAG-DGCR8 Only
Q8IU81	IRF2BP1	Interferon regulatory factor 2-binding protein 1	584	61.65	8.18	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	75.29	4.45	2.00	1.00	0.00	0.00	0.00	6.23	0.00	FLAG-DGCR8 Only
Q8IWS0	PHF6	PHD finger protein 6	365	41.26	8.68	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	38.96	6.85	2.00	1.00	0.00	0.00	0.00	5.28	0.00	FLAG-DGCR8 Only
Q92928	RAB1C	Putative Ras-related protein Rab-1C	201	22.00	5.43	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	89.74	22.89	4.00	1.00	0.00	0.00	0.00	6.49	0.00	FLAG-DGCR8 Only
Q9HAN9	NMNAT1	Nicotinamide/nicotinic acid mononucleotide adenylyltransferase 1	279	31.91	8.87	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	91.56	8.60	2.00	1.00	0.00	0.00	0.00	6.52	0.00	FLAG-DGCR8 Only
Q9UHB9	SRP68	Signal recognition particle subunit SRP68	627	70.69	8.56	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	90.95	5.10	3.00	1.00	0.00	0.00	0.00	6.51	0.00	FLAG-DGCR8 Only
Q9Y6X9	MORC2	MORC family CW-type zinc finger protein 2	1032	117.75	8.38	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	85.46	2.91	3.00	1.00	0.00	0.00	0.00	6.42	0.00	FLAG-DGCR8 Only
P60709	ACTB	Actin, cytoplasmic 1	375	41.71	5.48	1.00	0.00	0.00	1.00	0.00	0.00	99.40	13.33	4.00	216.45	18.67	6.00	93.75	9.33	4.00	0.00	1.12	-0.08	FLAG-DGCR8 Only
Q9UQE7	SMC3	Structural maintenance of chromosomes protein 3	1217	141.45	7.18	291.84	11.18	10.00	290.70	11.83	11.00	1.00	0.00	0.00	38.91	1.73	2.00	1.00	0.00	0.00	-0.01	5.28	0.00	FLAG-DGCR8 Only
Q7Z2W4	ZC3HAV1	Zinc finger CCCH-type antiviral protein 1	902	101.37	8.40	178.88	3.66	2.00	127.40	6.21	3.00	1.00	0.00	0.00	88.10	3.22	2.00	1.00	0.00	0.00	-0.49	6.46	0.00	FLAG-DGCR8 Only
P50991	CCT4	T-complex protein 1 subunit delta	539	57.89	7.83	431.46	31.54	13.00	271.75	11.50	4.00	59.37	4.27	3.00	124.31	5.75	3.00	95.13	8.16	5.00	-0.67	1.07	0.68	FLAG-DGCR8 Only
Q13263	TRIM28	Transcription intermediary factor 1-beta	835	88.49	5.77	408.61	25.39	14.00	229.03	14.37	8.00	1.00	0.00	0.00	229.54	11.26	6.00	1.00	0.00	0.00	-0.84	7.84	0.00	FLAG-DGCR8 Only
O14980	XPO1	Exportin-1	1071	123.31	6.06	760.80	28.01	24.00	383.21	14.85	12.00	1.00	0.00	0.00	98.84	3.92	3.00	1.00	0.00	0.00	-0.99	6.63	0.00	FLAG-DGCR8 Only
P14618	PKM	Pyruvate kinase PKM	531	57.90	7.84	500.96	37.29	15.00	217.48	10.73	4.00	1.00	0.00	0.00	131.34	16.38	6.00	1.00	0.00	0.00	-1.20	7.04	0.00	FLAG-DGCR8 Only
P51149	RAB7A	Ras-related protein Rab-7a	207	23.47	6.70	357.21	63.77	11.00	112.99	20.29	3.00	1.00	0.00	0.00	91.73	15.94	3.00	1.00	0.00	0.00	-1.66	6.52	0.00	FLAG-DGCR8 Only
P35250	RFC2	Replication factor C subunit 2	354	39.13	6.44	61.19	6.21	2.00	1.00	0.00	0.00	1.00	0.00	0.00	54.32	6.21	2.00	1.00	0.00	0.00	-5.94	5.76	0.00	FLAG-DGCR8 Only
Q14677	CLINT1	Clathrin interactor 1	625	68.22	6.42	81.22	4.00	2.00	1.00	0.00	0.00	1.00	0.00	0.00	65.68	3.52	2.00	1.00	0.00	0.00	-6.34	6.04	0.00	FLAG-DGCR8 Only

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
Q8IY67	RAVER1	Ribonucleoprotein PTB-binding 1	606	63.84	8.48	94.87	13.37	3.00	1.00	0.00	0.00	1.00	0.00	0.00	59.25	8.91	2.00	1.00	0.00	0.00	-6.57	5.89	0.00	FLAG-DGCR8 Only
P0CG48	UBC	Polyubiquitin-C	685	76.99	7.66	98.82	61.75	4.00	1.00	0.00	0.00	1.00	0.00	0.00	53.83	40.73	3.00	1.00	0.00	0.00	-6.63	5.75	0.00	FLAG-DGCR8 Only
P09874	PARP1	Poly [ADP-ribose] polymerase 1	1014	113.01	8.88	126.71	4.64	3.00	1.00	0.00	0.00	1.00	0.00	0.00	81.36	3.35	2.00	1.00	0.00	0.00	-6.99	6.35	0.00	FLAG-DGCR8 Only
P14868	DARS	Aspartate--tRNA ligase, cytoplasmic	501	57.10	6.55	175.58	20.36	9.00	1.00	0.00	0.00	1.00	0.00	0.00	39.26	3.39	2.00	1.00	0.00	0.00	-7.46	5.29	0.00	FLAG-DGCR8 Only
O60814	HIST1H2BK	Histone H2B type 1-K	126	13.88	10.32	188.21	35.71	3.00	1.00	0.00	0.00	1.00	0.00	0.00	247.21	35.71	1.00	1.00	0.00	0.00	-7.56	7.95	0.00	FLAG-DGCR8 Only
P11586	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic	935	101.50	7.30	316.52	9.52	8.00	1.00	0.00	0.00	1.00	0.00	0.00	126.88	5.67	5.00	1.00	0.00	0.00	-8.31	6.99	0.00	FLAG-DGCR8 Only
P78371	CCT2	T-complex protein 1 subunit beta	535	57.45	6.46	534.33	34.58	12.00	448.35	15.70	6.00	101.60	6.54	3.00	334.53	13.46	5.00	279.01	15.89	7.00	-0.25	1.72	1.46	FLAG-DGCR8/FLAG Drosha
P40227	CCT6A	T-complex protein 1 subunit zeta	531	57.99	6.68	360.19	22.79	10.00	127.12	8.29	4.00	1.00	0.00	0.00	77.19	4.71	2.00	32.31	3.58	2.00	-1.50	6.27	5.01	FLAG-DGCR8/FLAG Drosha
Q92841	DDX17	Probable ATP-dependent RNA helicase	729	80.22	8.27	486.34	21.26	7.00	467.11	21.54	9.00	1.00	0.00	0.00	148.57	8.09	2.00	246.53	13.58	5.00	-0.06	7.22	7.95	FLAG-DGCR8/FLAG Drosha
P17844	DDX5	Probable ATP-dependent RNA helicase	614	69.10	8.92	468.22	24.27	7.00	360.90	27.36	10.00	54.33	2.93	2.00	131.22	10.26	3.00	197.24	14.33	4.00	-0.38	1.27	1.86	FLAG-DGCR8/FLAG Drosha
O43143	DHX15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase	795	90.88	7.46	245.96	12.33	9.00	424.29	19.75	12.00	1.00	0.00	0.00	136.42	6.04	5.00	57.64	4.28	3.00	0.79	7.09	5.85	FLAG-DGCR8/FLAG Drosha
P31689	DNAJA1	DnaJ homolog subfamily A member 1	397	44.84	7.08	108.41	13.60	4.00	178.10	25.94	8.00	39.56	7.30	3.00	213.58	21.66	6.00	106.20	10.58	4.00	0.72	2.43	1.42	FLAG-DGCR8/FLAG Drosha
P25685	DNAJB1	DnaJ homolog subfamily B member 1	340	38.02	8.63	55.17	8.82	2.00	1.00	0.00	0.00	1.00	0.00	0.00	42.61	9.12	3.00	52.35	10.59	4.00	-5.79	5.41	5.71	FLAG-DGCR8/FLAG Drosha
Q15029	EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component	972	109.37	5.00	515.90	27.57	18.00	663.67	23.35	14.00	41.39	1.95	2.00	83.71	3.91	3.00	234.09	11.32	7.00	0.36	1.02	2.50	FLAG-DGCR8/FLAG Drosha
Q06787	FMR1	Fragile X mental retardation protein 1	632	71.13	7.42	146.29	6.33	3.00	108.11	4.91	2.00	1.00	0.00	0.00	44.79	2.53	2.00	43.99	2.53	2.00	-0.44	5.49	5.46	FLAG-DGCR8/FLAG Drosha
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	825	90.53	6.00	772.07	36.97	22.00	1065.73	39.39	28.00	132.15	12.36	6.00	606.18	32.12	22.00	410.04	24.85	13.00	0.47	2.20	1.63	FLAG-DGCR8/FLAG Drosha
Q9BUJ2	HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1	856	95.68	6.92	158.42	6.07	4.00	124.66	4.79	3.00	1.00	0.00	0.00	135.06	6.54	4.00	113.32	3.86	3.00	-0.35	7.08	6.82	FLAG-DGCR8/FLAG Drosha
O14654	IRS4	Insulin receptor substrate 4	1257	133.68	8.44	134.33	5.49	5.00	80.01	3.66	4.00	1.00	0.00	0.00	53.37	1.43	2.00	51.07	2.23	3.00	-0.75	5.74	5.67	FLAG-DGCR8/FLAG Drosha
Q14974	KPNB1	Importin subunit beta-1	876	97.11	4.78	483.06	22.72	14.00	428.90	13.93	9.00	1.00	0.00	0.00	64.10	3.08	2.00	92.30	3.08	2.00	-0.17	6.00	6.53	FLAG-DGCR8/FLAG Drosha
P43243	MATR3	Matrin-3	847	94.56	6.25	647.77	33.77	20.00	483.54	17.36	10.00	1.00	0.00	0.00	88.93	5.79	3.00	76.37	2.72	2.00	-0.42	6.47	6.25	FLAG-DGCR8/FLAG Drosha
Q9UMS4	PRPF19	Pre-mRNA-processing factor 19	504	55.15	6.61	300.36	35.32	8.00	326.87	54.96	13.00	1.00	0.00	0.00	93.11	20.04	6.00	62.22	11.31	4.00	0.12	6.54	5.96	FLAG-DGCR8/FLAG Drosha
P38159	RBMX	RNA-binding motif protein, X chromosome	391	42.31	10.05	286.65	27.11	11.00	300.07	21.99	9.00	1.00	0.00	0.00	72.78	10.74	4.00	67.12	7.93	3.00	0.07	6.19	6.07	FLAG-DGCR8/FLAG Drosha
Q9Y265	RUVBL1	RuvB-like 1	456	50.20	6.42	413.55	34.65	11.00	360.61	31.80	11.00	1.00	0.00	0.00	72.33	10.75	4.00	52.79	5.26	2.00	-0.20	6.18	5.72	FLAG-DGCR8/FLAG Drosha
Q9Y230	RUVBL2	RuvB-like 2	463	51.12	5.64	446.18	31.32	12.00	458.06	31.53	13.00	1.00	0.00	0.00	68.23	7.34	3.00	64.64	4.75	2.00	0.04	6.09	6.01	FLAG-DGCR8/FLAG Drosha
O75533	SF3B1	Splicing factor 3B subunit 1	1304	145.74	7.09	627.22	18.25	16.00	556.59	14.34	14.00	1.00	0.00	0.00	44.63	1.92	2.00	70.97	2.84	3.00	-0.17	5.48	6.15	FLAG-DGCR8/FLAG Drosha
Q13148	TARDBP	TAR DNA-binding protein 43	414	44.71	6.19	471.38	32.13	12.00	329.74	20.77	7.00	1.00	0.00	0.00	253.42	15.22	5.00	122.41	12.32	3.00	-0.52	7.99	6.94	FLAG-DGCR8/FLAG Drosha
P17987	TCP1	T-complex protein 1 subunit alpha	556	60.31	6.11	352.39	24.46	11.00	1.00	0.00	0.00	1.00	0.00	0.00	113.05	9.89	4.00	62.87	3.96	2.00	-8.46	6.82	5.97	FLAG-DGCR8/FLAG Drosha
P04350	TUBB4A	Tubulin beta-4A chain	444	49.55	4.88	1484.51	69.59	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1013.02	50.90	1.00	645.06	52.03	1.00	-10.54	9.98	9.33	FLAG-DGCR8/FLAG Drosha
P12956	XRCC6	X-ray repair cross-complementing protein 6	609	69.80	6.64	123.15	13.46	6.00	61.92	5.42	3.00	29.36	4.27	2.00	95.16	7.39	4.00	73.36	3.12	2.00	-0.99	1.70	1.32	FLAG-DGCR8/FLAG Drosha
P62913	RPL11	60S ribosomal protein L11	178	20.24	9.60	139.39	19.66	3.00	322.07	54.49	10.00	147.86	19.66	3.00	175.26	28.09	5.00	111.16	19.66	3.00	1.21	0.25	-0.41	T7-DGCR8 Only
P30050	RPL12	60S ribosomal protein L12	165	17.81	9.42	138.01	33.94	4.00	318.80	44.85	5.00	190.41	33.94	4.00	378.81	54.55	6.00	88.52	24.85	3.00	1.21	0.99	-1.11	T7-DGCR8 Only
Q14690	PDCD11	Protein RRP5 homolog	1871	208.57	8.87	57.18	1.76	3.00	310.30	6.41	10.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	2.44	0.00	0.00	T7-DGCR8 Only
Q12788	TBL3	Transducin beta-like protein 3	808	88.98	6.90	81.59	4.70	3.00	358.45	16.46	10.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	2.14	0.00	0.00	T7-DGCR8 Only
Q07955	SRSF1	Serine/arginine-rich splicing factor 1	248	27.73	10.36	47.62	16.94	3.00	170.85	20.56	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.84	0.00	0.00	T7-DGCR8 Only
Q9Y3T9	NOC2L	Nucleolar complex protein 2 homolog	749	84.87	5.62	83.88	3.74	3.00	259.35	11.62	8.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.63	0.00	0.00	T7-DGCR8 Only
O75319	DUSP11	RNA/RNP complex-1-interacting phosphatase	330	38.91	9.41	52.83	9.70	2.00	148.92	13.33	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.50	0.00	0.00	T7-DGCR8 Only
P26368	U2AF2	Splicing factor U2AF 65 kDa subunit	475	53.47	9.09	235.88	28.00	8.00	631.01	47.58	13.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.42	0.00	0.00	T7-DGCR8 Only
Q9Y2R4	DDX52	Probable ATP-dependent RNA helicase DDX52	599	67.46	9.67	80.37	7.18	3.00	211.62	10.68	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.40	0.00	0.00	T7-DGCR8 Only
Q5BKZ1	ZNF326	DBIRD complex subunit ZNF326	582	65.61	5.15	83.77	6.70	3.00	216.94	13.75	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.37	0.00	0.00	T7-DGCR8 Only
P18583	SON	Protein SON	2426	263.66	5.64	45.09	0.82	2.00	107.58	1.69	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.25	0.00	0.00	T7-DGCR8 Only
Q13243	SRSF5	Serine/arginine-rich splicing factor 5	272	31.25	11.59	92.56	8.82	1.00	219.81	25.00	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.25	0.00	0.00	T7-DGCR8 Only
P28799	GRN	Granulins	593	63.50	6.83	85.85	4.55	2.00	197.35	9.95	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.20	0.00	0.00	T7-DGCR8 Only
Q9Y3Y2	CHTOP	Chromatin target of PRMT1 protein	248	26.38	12.23	133.23	14.52	4.00	289.58	32.26	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.12	0.00	0.00	T7-DGCR8 Only
Q13283	G3BP1	Ras GTPase-activating protein-binding protein 1	466	52.13	5.52	78.38	4.29	2.00	168.96	7.94	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.11	0.00	0.00	T7-DGCR8 Only
Q13428	TCOF1	Treacle protein	1488	152.02	9.04	62.44	1.41	2.00	134.45	1.95	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.11	0.00	0.00	T7-DGCR8 Only
Q14839	CHD4	Chromodomain-helicase-DNA-binding protein 4	1912	217.87	5.86	185.42	4.45	6.00	387.47	7.27	11.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.06	0.00	0.00	T7-DGCR8 Only
Q15427	SF3B4	Splicing factor 3B subunit 4	424	44.36	8.56	95.23	13.44	3.00	195.73	23.11	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.04	0.00	0.00	T7-DGCR8 Only
O00567	NOP56	Nucleolar protein 56	594	66.01	9.19	175.80	17.17	6.00	359.01	26.60	12.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.03	0.00	0.00	T7-DGCR8 Only
Q9BQ39	DDX50	ATP-dependent RNA helicase DDX50	737	82.51	9.17	1.00	0.00	0.00	635.38	29.04	15.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	9.31	0.00	0.00	T7-DGCR8 Only
Q8IY81	FTSJ3	pre-rRNA processing protein FTSJ3	847	96.50	8.40	1.00	0.00	0.00	603.83	23.38	14.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	9.24	0.00	0.00	T7-DGCR8 Only
O15381	NVL	Nuclear valosin-containing protein-like	856	94.99	6.48	1.00	0.00	0.00	585.94	18.93	12.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	9.19	0.00	0.00	T7-DGCR8 Only
Q9H0A0	NAT10	N-acetyltransferase 10	1025	115.66	8.27	1.00	0.00	0.00	563.17	22.63	16.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	9.14	0.00	0.00	T7-DGCR8 Only
O94813	SLIT2	Slit homolog 2 protein	1529	169.76	7.06	1.00	0.00	0.00	543.18	11.25	11.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	9.09	0.00	0.00	T7-DGCR8 Only
Q8TDD1	DDX54	ATP-dependent RNA helicase DDX54	881	98.53	10.02	1.00	0.00	0.00	459.62	20.20	13.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.84	0.00	0.00	T7-DGCR8 Only
Q9NVP1	DDX18	ATP-dependent RNA helicase DDX18	670	75.36	9.50	1.00	0.00	0.00	386.31	23.28	12.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.59	0.00	0.00	T7-DGCR8 Only
Q15050	RRS1	Ribosome biogenesis regulatory protein homolog	365	41.17	10.70	1.00	0.00	0.00	380.18	35.34	13.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.57	0.00	0.00	T7-DGCR8 Only
Q96KR1	ZFR	Zinc finger RNA-binding protein	1074	116.94	9.04	1.00	0.00	0.00	343.56	14.34	8.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.42	0.00	0.00	T7-DGCR8 Only
Q99575	POP1	Ribonucleases P/MRP protein subunit POP1	1024	114.64	9.22	1.00	0.00	0.00	337.76	12.89	9.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.40	0.00	0.00	T7-DGCR8 Only
Q9HCE1	MOV10	Putative helicase MOV-10	1003	113.60	8.82	1.00	0.00	0.00	334.68	10.07	8.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.39	0.00	0.00	T7-DGCR8 Only
Q9GZR7	DDX24	ATP-dependent RNA helicase DDX24	859	96.27	9.06	1.00	0.00	0.00	326.08	16.53	10.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.35	0.00	0.00	T7-DGCR8 Only
O00541	PES1	Pescadillo homolog	588	67.96	7.33	1.00	0.00	0.00	320.57	23.13	15.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.32	0.00	0.00	T7-DGCR8 Only
Q9NQ55	PPAN	Suppressor of SWI4 1 homolog	473	53.16	10.13	1.00	0.00	0.00	315.89	23.68	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.30	0.00	0.00	T7-DGCR8 Only
Q14692	BMS1	Ribosome biogenesis protein BMS1 homolog	1282	145.72	6.44	1.00	0.00	0.00	312.97	8.11	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.29	0.00	0.00	T7-DGCR8 Only
Q5JTH9	RRP12	RRP12-like protein	1297	143.61	8.75	1.00	0.00	0.00	309.84	6.71	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.28	0.00	0.00	T7-DGCR8 Only
Q9Y2P8	RCL1	RNA 3'-terminal phosphate cyclase-like protein	373	40.82	9.26	1.00	0.00	0.00	295.42	27.88	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.21	0.00	0.00	T7-DGCR8 Only

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
Q14137	BOP1	Ribosome biogenesis protein BOP1	746	83.58	6.19	1.00	0.00	0.00	286.59	10.59	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.16	0.00	0.00	T7-DGCR8 Only
Q96GQ7	DDX27	Probable ATP-dependent RNA helicase DDX27	796	89.78	9.28	1.00	0.00	0.00	271.14	13.57	10.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.08	0.00	0.00	T7-DGCR8 Only
Q8IY37	DHX37	Probable ATP-dependent RNA helicase DHX37	1157	129.46	8.10	1.00	0.00	0.00	260.79	7.26	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.03	0.00	0.00	T7-DGCR8 Only
Q13823	GNL2	Nucleolar GTP-binding protein 2	731	83.60	9.25	1.00	0.00	0.00	258.90	16.83	9.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.02	0.00	0.00	T7-DGCR8 Only
Q15397	KIAA0020	Pumilio domain-containing protein KIAA0020	648	73.54	9.64	1.00	0.00	0.00	254.35	17.28	8.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.99	0.00	0.00	T7-DGCR8 Only
Q9NXF1	TEX10	Testis-expressed sequence 10 protein	929	105.61	9.36	1.00	0.00	0.00	247.47	8.50	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.95	0.00	0.00	T7-DGCR8 Only
Q9H6R4	NOL6	Nucleolar protein 6	1146	127.51	7.64	1.00	0.00	0.00	246.77	5.50	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.95	0.00	0.00	T7-DGCR8 Only
Q15269	PWP2	Periodic tryptophan protein 2 homolog	919	102.39	6.15	1.00	0.00	0.00	236.35	8.92	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.88	0.00	0.00	T7-DGCR8 Only
Q8WTT2	NOC3L	Nucleolar complex protein 3 homolog	800	92.49	9.17	1.00	0.00	0.00	233.11	13.25	8.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.86	0.00	0.00	T7-DGCR8 Only
Q12789	GTF3C1	General transcription factor 3C polypeptide 1	2109	238.72	7.30	1.00	0.00	0.00	227.65	5.93	9.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.83	0.00	0.00	T7-DGCR8 Only
Q9NQT4	EXOSC5	Exosome complex component RRP46	235	25.23	7.59	1.00	0.00	0.00	213.61	34.89	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.74	0.00	0.00	T7-DGCR8 Only
Q13610	PWP1	Periodic tryptophan protein 1 homolog	501	55.79	4.77	1.00	0.00	0.00	211.18	23.75	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.72	0.00	0.00	T7-DGCR8 Only
Q9GZL7	WDR12	Ribosome biogenesis protein WDR12	423	47.68	5.90	1.00	0.00	0.00	204.97	14.18	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.68	0.00	0.00	T7-DGCR8 Only
Q9H0S4	DDX47	Probable ATP-dependent RNA helicase DDX47	455	50.61	9.10	1.00	0.00	0.00	203.30	13.41	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.67	0.00	0.00	T7-DGCR8 Only
Q9Y221	NIP7	60S ribosome subunit biogenesis protein NIP7 homolog	180	20.45	8.51	1.00	0.00	0.00	185.42	40.00	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.53	0.00	0.00	T7-DGCR8 Only
Q8N8A6	DDX51	ATP-dependent RNA helicase DDX51	666	72.41	8.16	1.00	0.00	0.00	177.25	11.26	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.47	0.00	0.00	T7-DGCR8 Only
O60832	DKC1	H/ACA ribonucleoprotein complex subunit 4	514	57.64	9.42	1.00	0.00	0.00	174.38	15.95	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.45	0.00	0.00	T7-DGCR8 Only
Q9NZM5	GLTSCR2	Glioma tumor suppressor candidate region gene 2 protein	478	54.36	10.32	1.00	0.00	0.00	168.83	6.49	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.40	0.00	0.00	T7-DGCR8 Only
O00411	POLRMT	DNA-directed RNA polymerase, mitochondrial	1230	138.53	8.98	1.00	0.00	0.00	167.44	7.56	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.39	0.00	0.00	T7-DGCR8 Only
Q9NQZ2	UTP3	Something about silencing protein 10	479	54.53	5.62	1.00	0.00	0.00	163.26	8.98	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.35	0.00	0.00	T7-DGCR8 Only
Q9BVJ6	UTP14A	U3 small nucleolar RNA-associated protein 14 homolog A	771	87.92	7.87	1.00	0.00	0.00	162.45	6.74	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.34	0.00	0.00	T7-DGCR8 Only
Q15024	EXOSC7	Exosome complex component RRP42	291	31.80	5.19	1.00	0.00	0.00	156.34	24.05	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.29	0.00	0.00	T7-DGCR8 Only
Q9Y3B2	EXOSC1	Exosome complex component CSL4	195	21.44	8.24	1.00	0.00	0.00	154.09	21.03	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.27	0.00	0.00	T7-DGCR8 Only
Q5T3I0	GPATCH4	G patch domain-containing protein 4	446	50.35	9.63	1.00	0.00	0.00	153.93	12.56	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.27	0.00	0.00	T7-DGCR8 Only
Q13601	KRR1	KRR1 small subunit processome component homolog	381	43.64	9.77	1.00	0.00	0.00	152.03	9.97	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.25	0.00	0.00	T7-DGCR8 Only
Q9NY93	DDX56	Probable ATP-dependent RNA helicase DDX56	547	61.55	9.26	1.00	0.00	0.00	148.77	10.79	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.22	0.00	0.00	T7-DGCR8 Only
P78362	SRPK2	SRSF protein kinase 2	688	77.48	4.97	1.00	0.00	0.00	147.04	7.85	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.20	0.00	0.00	T7-DGCR8 Only
Q05519	SRSF11	Serine/arginine-rich splicing factor 11	484	53.51	10.52	1.00	0.00	0.00	145.10	9.50	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.18	0.00	0.00	T7-DGCR8 Only
Q8WUA4	GTF3C2	General transcription factor 3C polypeptide 2	911	100.62	7.31	1.00	0.00	0.00	145.09	6.37	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.18	0.00	0.00	T7-DGCR8 Only
Q8NHQ9	DDX55	ATP-dependent RNA helicase DDX55	600	68.50	9.25	1.00	0.00	0.00	141.23	8.50	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.14	0.00	0.00	T7-DGCR8 Only
Q9Y3C1	NOP16	Nucleolar protein 16	178	21.18	9.94	1.00	0.00	0.00	140.74	26.97	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.14	0.00	0.00	T7-DGCR8 Only
Q9H8H2	DDX31	Probable ATP-dependent RNA helicase DDX31	851	94.03	9.99	1.00	0.00	0.00	139.50	5.52	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.12	0.00	0.00	T7-DGCR8 Only
P62491	RAB11A	Ras-related protein Rab-11A	216	24.38	6.57	1.00	0.00	0.00	137.61	21.76	4.00	62.27	10.19	2.00	56.03	10.19	2.00	49.39	10.19	2.00	7.10	-0.15	-0.33	T7-DGCR8 Only
Q96SI9	STRBP	Spermatid perinuclear RNA-binding protein	672	73.61	8.72	1.00	0.00	0.00	136.37	7.89	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.09	0.00	0.00	T7-DGCR8 Only
Q9H5H4	ZNF768	Zinc finger protein 768	540	60.19	7.97	1.00	0.00	0.00	133.07	10.00	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.06	0.00	0.00	T7-DGCR8 Only
O95602	POLR1A	DNA-directed RNA polymerase I subunit RPA1	1720	194.69	7.03	1.00	0.00	0.00	129.75	4.30	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.02	0.00	0.00	T7-DGCR8 Only
P42285	SKIV2L2	Superkiller viralicidic activity 2-like 2	1042	117.73	6.52	1.00	0.00	0.00	126.17	5.57	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.98	0.00	0.00	T7-DGCR8 Only
O60684	KPNA6	Importin subunit alpha-7	536	59.99	4.98	1.00	0.00	0.00	123.98	7.28	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.95	0.00	0.00	T7-DGCR8 Only
Q9NWT1	PAK1IP1	p21-activated protein kinase-interacting protein 1	392	43.94	8.91	1.00	0.00	0.00	122.90	9.44	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.94	0.00	0.00	T7-DGCR8 Only
Q96B26	EXOSC8	Exosome complex component RRP43	276	30.02	5.30	1.00	0.00	0.00	119.12	18.48	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.90	0.00	0.00	T7-DGCR8 Only
Q5T8P6	RBM26	RNA-binding protein 26	1007	113.53	9.16	1.00	0.00	0.00	118.69	3.77	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.89	0.00	0.00	T7-DGCR8 Only
Q9H3G5	CPVL	Probable serine carboxypeptidase CPVL	476	54.13	5.62	1.00	0.00	0.00	114.35	7.77	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.84	0.00	0.00	T7-DGCR8 Only
Q03701	CEBPZ	CCAAT/enhancer-binding protein zeta	1054	120.90	5.94	1.00	0.00	0.00	112.47	3.04	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.81	0.00	0.00	T7-DGCR8 Only
Q5C9Z4	NOM1	Nucleolar MIF4G domain-containing protein 1	860	96.20	8.10	1.00	0.00	0.00	112.40	6.05	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.81	0.00	0.00	T7-DGCR8 Only
Q9NY61	AATF	Protein AATF	560	63.09	4.94	1.00	0.00	0.00	111.80	11.43	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.80	0.00	0.00	T7-DGCR8 Only
Q86U38	NOP9	Nucleolar protein 9	636	69.39	7.28	1.00	0.00	0.00	111.30	9.43	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.80	0.00	0.00	T7-DGCR8 Only
Q13523	PRPF4B	Serine/threonine-protein kinase PRP4 homolog	1007	116.92	10.26	1.00	0.00	0.00	109.57	2.78	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.78	0.00	0.00	T7-DGCR8 Only
Q9UKD2	MRTO4	mRNA turnover protein 4 homolog	239	27.54	8.29	1.00	0.00	0.00	109.17	18.41	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.77	0.00	0.00	T7-DGCR8 Only
P78316	NOP14	Nucleolar protein 14	857	97.61	7.58	1.00	0.00	0.00	106.95	4.43	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.74	0.00	0.00	T7-DGCR8 Only
P19388	POLR2E	DNA-directed RNA polymerases I, II, and III subunit RPABC1	210	24.54	5.95	1.00	0.00	0.00	105.45	17.14	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.72	0.00	0.00	T7-DGCR8 Only
O15226	NKRF	NF-kappa-B-repressing factor	690	77.62	8.79	1.00	0.00	0.00	105.13	5.94	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.72	0.00	0.00	T7-DGCR8 Only
Q14146	URB2	Unhealthy ribosome biogenesis protein 2 homolog	1524	170.43	7.31	1.00	0.00	0.00	104.54	3.28	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.71	0.00	0.00	T7-DGCR8 Only
Q9H583	HEATR1	HEAT repeat-containing protein 1	2144	242.22	6.54	1.00	0.00	0.00	103.00	1.63	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.69	0.00	0.00	T7-DGCR8 Only
Q9Y5J1	UTP18	U3 small nucleolar RNA-associated protein 18 homolog	556	61.96	8.76	1.00	0.00	0.00	100.97	8.99	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.66	0.00	0.00	T7-DGCR8 Only
Q5T280	C9orf114	Uncharacterized protein C9orf114	376	41.98	7.43	1.00	0.00	0.00	98.53	10.37	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.62	0.00	0.00	T7-DGCR8 Only
Q9UNX4	WDR3	WD repeat-containing protein 3	943	106.03	6.64	1.00	0.00	0.00	95.55	2.23	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.58	0.00	0.00	T7-DGCR8 Only
O43818	RRP9	U3 small nucleolar RNA-interacting protein 2	475	51.81	7.85	1.00	0.00	0.00	95.37	6.11	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.58	0.00	0.00	T7-DGCR8 Only
P32322	PYCR1	Pyrroline-5-carboxylate reductase 1, mitochondrial	319	33.34	7.61	1.00	0.00	0.00	94.99	7.52	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.57	0.00	0.00	T7-DGCR8 Only
Q2NL82	TSR1	Pre-rRNA-processing protein TSR1 homolog	804	91.75	7.42	1.00	0.00	0.00	94.58	3.36	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.56	0.00	0.00	T7-DGCR8 Only
Q96G21	IMP4	U3 small nucleolar ribonucleoprotein protein IMP4	291	33.74	9.47	1.00	0.00	0.00	94.56	11.68	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.56	0.00	0.00	T7-DGCR8 Only
Q9Y388	RBMX2	RNA-binding motif protein, X-linked 2	322	37.31	9.83	1.00	0.00	0.00	94.18	13.35	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.56	0.00	0.00	T7-DGCR8 Only
Q13206	DDX10	Probable ATP-dependent RNA helicase DDX10	875	100.83	8.63	1.00	0.00	0.00	94.05	5.94	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.56	0.00	0.00	T7-DGCR8 Only
Q9Y3B9	RRP15	RRP15-like protein	282	31.46	5.52	1.00	0.00	0.00	93.87	14.89	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.55	0.00	0.00	T7-DGCR8 Only
O43172	PRPF4	U4/U6 small nuclear ribonucleoprotein Prp4	522	58.41	7.42	1.00	0.00	0.00	91.62	5.75	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.52	0.00	0.00	T7-DGCR8 Only
Q9H9Y2	RPF1	Ribosome production factor 1	349	40.09	10.01	1.00	0.00	0.00	91.15	10.60	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.51	0.00	0.00	T7-DGCR8 Only

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
A4D1E9	GTPBP10	GTP-binding protein 10	387	42.91	9.03	1.00	0.00	0.00	90.58	10.85	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.50	0.00	0.00	T7-DGCR8 Only
Q8NE71	ABCF1	ATP-binding cassette sub-family F member 1	845	95.87	6.80	1.00	0.00	0.00	87.90	2.72	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.46	0.00	0.00	T7-DGCR8 Only
P15927	RPA2	Replication protein A 32 kDa subunit	270	29.23	6.15	1.00	0.00	0.00	87.21	16.67	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.45	0.00	0.00	T7-DGCR8 Only
P35611	ADD1	Alpha-adducin	737	80.90	5.83	1.00	0.00	0.00	86.18	3.80	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.43	0.00	0.00	T7-DGCR8 Only
Q14669	TRIP12	E3 ubiquitin-protein ligase TRIP12	1992	220.30	8.48	1.00	0.00	0.00	84.25	1.20	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.40	0.00	0.00	T7-DGCR8 Only
P57678	GEMIN4	Gem-associated protein 4	1058	119.96	6.04	1.00	0.00	0.00	83.04	1.89	1.00	193.74	8.79	8.00	279.31	11.44	8.00	352.41	13.80	12.00	6.38	0.53	0.86	T7-DGCR8 Only
Q5PRF9	SAMD4B	Protein Smaug homolog 2	694	75.44	6.83	1.00	0.00	0.00	81.53	4.03	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.35	0.00	0.00	T7-DGCR8 Only
Q9BRT6	LLPH	Protein LLP homolog	129	15.22	10.37	1.00	0.00	0.00	81.02	10.85	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.34	0.00	0.00	T7-DGCR8 Only
P54709	ATP1B3	Sodium/potassium-transporting ATPase subunit beta-3	279	31.49	8.35	1.00	0.00	0.00	79.69	8.96	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.32	0.00	0.00	T7-DGCR8 Only
Q01130	SRSF2	Serine/arginine-rich splicing factor 2	221	25.46	11.85	1.00	0.00	0.00	77.79	14.03	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.28	0.00	0.00	T7-DGCR8 Only
Q9BXY0	MAK16	Protein MAK16 homolog	300	35.35	5.38	1.00	0.00	0.00	76.63	9.00	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.26	0.00	0.00	T7-DGCR8 Only
Q96QD9	FYTTD1	UAP56-interacting factor	318	35.80	11.78	1.00	0.00	0.00	76.01	7.55	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.25	0.00	0.00	T7-DGCR8 Only
Q12874	SF3A3	Splicing factor 3A subunit 3	501	58.81	5.38	1.00	0.00	0.00	75.33	4.99	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.24	0.00	0.00	T7-DGCR8 Only
Q9H7E9	C8orf33	UPF0488 protein C8orf33	229	24.98	9.95	1.00	0.00	0.00	74.92	15.28	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.23	0.00	0.00	T7-DGCR8 Only
O95104	SCAF4	Splicing factor, arginine/serine-rich 15	1147	125.79	9.55	1.00	0.00	0.00	74.83	2.44	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.23	0.00	0.00	T7-DGCR8 Only
E9PRG8	C11orf98	Uncharacterized protein C11orf98	122	13.79	11.69	1.00	0.00	0.00	74.82	15.57	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.23	0.00	0.00	T7-DGCR8 Only
O75683	SURF6	Surfeit locus protein 6	361	41.43	10.64	1.00	0.00	0.00	74.32	12.74	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.22	0.00	0.00	T7-DGCR8 Only
Q8NFW8	CMAS	N-acylneuraminate cytidylyltransferase	434	48.35	7.93	1.00	0.00	0.00	74.30	8.76	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.22	0.00	0.00	T7-DGCR8 Only
O95478	NSA2	Ribosome biogenesis protein NSA2 homolog	260	30.05	10.27	1.00	0.00	0.00	74.05	10.77	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.21	0.00	0.00	T7-DGCR8 Only
P78346	RPP30	Ribonuclease P protein subunit p30	268	29.30	8.91	1.00	0.00	0.00	73.43	7.09	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.20	0.00	0.00	T7-DGCR8 Only
Q9HCM4	EPB41L5	Band 4.1-like protein 5	733	81.80	6.58	1.00	0.00	0.00	72.76	3.00	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.19	0.00	0.00	T7-DGCR8 Only
Q5M9Q1	NKAPL	NKAP-like protein	402	46.28	9.72	1.00	0.00	0.00	72.48	5.22	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.18	0.00	0.00	T7-DGCR8 Only
O60488	ACSL4	Long-chain-fatty-acid--CoA ligase 4	711	79.14	8.38	1.00	0.00	0.00	72.24	4.36	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.17	0.00	0.00	T7-DGCR8 Only
Q9Y5V3	MAGED1	Melanoma-associated antigen D1	778	86.11	5.83	1.00	0.00	0.00	71.69	2.96	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.16	0.00	0.00	T7-DGCR8 Only
Q96DH6	MSI2	RNA-binding protein Musashi homolog 2	328	35.17	8.48	1.00	0.00	0.00	70.90	9.15	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.15	0.00	0.00	T7-DGCR8 Only
A6NHQ2	FBLL1	rRNA/tRNA 2'-O-methyltransferase fibrillarin-like protein 1	333	34.65	10.33	1.00	0.00	0.00	66.46	9.61	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.05	0.00	0.00	T7-DGCR8 Only
O75367	H2AFY	Core histone macro-H2A.1	372	39.59	9.79	1.00	0.00	0.00	66.03	16.40	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.04	0.00	0.00	T7-DGCR8 Only
Q8NDF8	PAPD5	Non-canonical poly(A) RNA polymerase PAPD5	572	63.23	8.98	1.00	0.00	0.00	65.56	4.55	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.03	0.00	0.00	T7-DGCR8 Only
P51116	FXR2	Fragile X mental retardation syndrome-related protein 2	673	74.18	6.23	1.00	0.00	0.00	64.92	7.43	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.02	0.00	0.00	T7-DGCR8 Only
Q86U86	PBRM1	Protein polybromo-1	1689	192.83	6.89	1.00	0.00	0.00	64.57	1.30	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.01	0.00	0.00	T7-DGCR8 Only
Q8WVM0	TFB1M	Dimethyladenosine transferase 1, mitochondrial	346	39.52	9.26	1.00	0.00	0.00	63.87	9.54	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.00	0.00	0.00	T7-DGCR8 Only
Q5UIP0	RIF1	Telomere-associated protein RIF1	2472	274.29	5.52	1.00	0.00	0.00	63.55	1.17	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.99	0.00	0.00	T7-DGCR8 Only
Q9UII4	HERC5	E3 ISG15--protein ligase HERC5	1024	116.78	7.65	1.00	0.00	0.00	59.81	2.05	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.90	0.00	0.00	T7-DGCR8 Only
Q08J23	NSUN2	tRNA (cytosine(34)-C(5))-methyltransferase	767	86.42	6.77	1.00	0.00	0.00	59.69	2.74	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.90	0.00	0.00	T7-DGCR8 Only
Q86YP4	GATAD2A	Transcriptional repressor p66-alpha	633	68.02	9.94	1.00	0.00	0.00	59.46	3.63	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.89	0.00	0.00	T7-DGCR8 Only
O75691	UTP20	Small subunit processome component 20 homolog	2785	318.18	7.39	1.00	0.00	0.00	59.36	0.79	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.89	0.00	0.00	T7-DGCR8 Only
P20674	COX5A	Cytochrome c oxidase subunit 5A, mitochondrial	150	16.75	6.79	1.00	0.00	0.00	58.38	10.67	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.87	0.00	0.00	T7-DGCR8 Only
Q9NY12	GAR1	H/ACA ribonucleoprotein complex subunit 1	217	22.33	10.92	1.00	0.00	0.00	57.16	12.44	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.84	0.00	0.00	T7-DGCR8 Only
Q9UNF1	MAGED2	Melanoma-associated antigen D2	606	64.91	9.32	1.00	0.00	0.00	56.76	3.14	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.83	0.00	0.00	T7-DGCR8 Only
Q6PKG0	LARP1	La-related protein 1	1096	123.43	8.82	1.00	0.00	0.00	56.58	1.82	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.82	0.00	0.00	T7-DGCR8 Only
Q6PK04	CCDC137	Coiled-coil domain-containing protein 137	289	33.21	10.93	1.00	0.00	0.00	53.51	7.61	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.74	0.00	0.00	T7-DGCR8 Only
Q9H8H0	NOL11	Nucleolar protein 11	719	81.07	6.07	1.00	0.00	0.00	51.26	4.59	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.68	0.00	0.00	T7-DGCR8 Only
Q9UN86	G3BP2	Ras GTPase-activating protein-binding protein 2	482	54.09	5.55	1.00	0.00	0.00	49.76	6.22	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.64	0.00	0.00	T7-DGCR8 Only
Q9NVN8	GNL3L	Guanine nucleotide-binding protein-like 3-like protein	582	65.53	8.44	1.00	0.00	0.00	49.19	3.26	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.62	0.00	0.00	T7-DGCR8 Only
Q15046	KARS	Lysine--tRNA ligase	597	68.00	6.35	1.00	0.00	0.00	49.19	3.02	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.62	0.00	0.00	T7-DGCR8 Only
Q8NB90	SPATA5	Spermatogenesis-associated protein 5	893	97.84	5.66	1.00	0.00	0.00	47.50	2.58	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.57	0.00	0.00	T7-DGCR8 Only
Q9Y305	ACOT9	Acyl-coenzyme A thioesterase 9, mitochondrial	439	49.87	8.60	1.00	0.00	0.00	46.85	4.33	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.55	0.00	0.00	T7-DGCR8 Only
Q9H9Y6	POLR1B	DNA-directed RNA polymerase I subunit RPA2	1135	128.15	7.83	1.00	0.00	0.00	45.24	2.20	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.50	0.00	0.00	T7-DGCR8 Only
Q99547	MPHOSPH6	M-phase phosphoprotein 6	160	19.01	5.20	1.00	0.00	0.00	44.82	17.50	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.49	0.00	0.00	T7-DGCR8 Only
O60683	PEX10	Peroxisome biogenesis factor 10	326	37.05	9.99	1.00	0.00	0.00	43.35	3.37	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.44	0.00	0.00	T7-DGCR8 Only
Q7Z7L1	SLFN11	Schlafen family member 11	901	102.77	7.77	1.00	0.00	0.00	39.16	0.78	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.29	0.00	0.00	T7-DGCR8 Only
Q15643	TRIP11	Thyroid receptor-interacting protein 11	1979	227.45	5.26	1.00	0.00	0.00	31.17	0.61	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	4.96	0.00	0.00	T7-DGCR8 Only
P61619	SEC61A1	Protein transport protein Sec61 subunit alpha isoform 1	476	52.23	8.06	1.00	0.00	0.00	30.89	6.51	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	4.95	0.00	0.00	T7-DGCR8 Only
Q14684	RRP1B	Ribosomal RNA processing protein 1 homolog B	758	84.38	9.76	1.00	0.00	0.00	272.48	11.48	7.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	8.09	0.00	0.00	T7-DGCR8 Only
O60287	URB1	Nucleolar pre-ribosomal-associated protein 1	2271	254.23	6.47	1.00	0.00	0.00	208.91	3.08	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.71	0.00	0.00	T7-DGCR8 Only
O43159	RRP8	Ribosomal RNA-processing protein 8	456	50.68	9.42	1.00	0.00	0.00	156.16	6.58	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.29	0.00	0.00	T7-DGCR8 Only
P51398	DAP3	28S ribosomal protein S29, mitochondrial	398	45.54	8.88	1.00	0.00	0.00	152.67	13.32	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.25	0.00	0.00	T7-DGCR8 Only
Q9Y3A4	RRP7A	Ribosomal RNA-processing protein 7 homolog A	280	32.31	9.58	1.00	0.00	0.00	82.59	8.21	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.37	0.00	0.00	T7-DGCR8 Only
Q9UJV9	DDX41	Probable ATP-dependent RNA helicase DDX41	622	69.79	6.84	1.00	0.00	0.00	108.41	11.25	6.00	1.00	0.00	0.00	1.00	0.00	0.00	43.88	4.34	2.00	6.76	0.00	5.46	T7-DGCR8/FLAG-Drosha
P09651	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	372	38.72	9.13	410.20	31.72	8.00	845.78	47.31	14.00	118.83	7.53	2.00	189.06	12.90	4.00	304.63	17.20	5.00	1.04	0.67	1.36	T7-DGCR8/FLAG-Drosha
P22626	HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	353	37.41	8.95	210.06	24.65	7.00	452.58	35.69	12.00	1.00	0.00	0.00	1.00	0.00	0.00	155.63	15.01	4.00	1.11	0.00	7.28	T7-DGCR8/FLAG-Drosha
O60506	SYNCRIP	Heterogeneous nuclear ribonucleoprotein Q	623	69.56	8.59	1.00	0.00	0.00	180.23	14.45	4.00	1.00	0.00	0.00	1.00	0.00	0.00	81.65	4.65	2.00	7.49	0.00	6.35	T7-DGCR8/FLAG-Drosha
Q9BVA1	TUBB2B	Tubulin beta-2B chain	445	49.92	4.89	1.00	0.00	0.00	1512.83	61.57	2.00	1.00	0.00	0.00	1.00	0.00	0.00	738.53	53.03	2.00	10.56	0.00	9.53	T7-DGCR8/FLAG-Drosha
P61513	RPL37A	60S ribosomal protein L37a	92	10.27	10.43	1.00	0.00	0.00	332.28	42.39	3.00	1.00	0.00	0.00	122.63	32.61	2.00	1.00	0.00	0.00	8.38	6.94	0.00	Ribosomal Protein
P46777	RPL5	60S ribosomal protein L5	297	34.34	9.72	1.00	0.00	0.00	329.37	38.38	7.00	1.00	0.00	0.00	122.60	20.54	4.00	1.00	0.00	0.00	8.36	6.94	0.00	Ribosomal Protein

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
P35268	RPL22	60S ribosomal protein L22	128	14.78	9.19	1.00	0.00	0.00	302.94	53.13	6.00	84.90	18.75	2.00	204.87	41.41	5.00	74.55	18.75	2.00	8.24	1.27	-0.19	Ribosomal Protein
P50914	RPL14	60S ribosomal protein L14	215	23.42	10.93	1.00	0.00	0.00	256.09	19.53	5.00	1.00	0.00	0.00	156.39	19.07	4.00	77.69	16.28	3.00	8.00	7.29	6.28	Ribosomal Protein
P62910	RPL32	60S ribosomal protein L32	135	15.85	11.33	1.00	0.00	0.00	228.37	42.22	7.00	1.00	0.00	0.00	138.43	29.63	4.00	1.00	0.00	0.00	7.84	7.11	0.00	Ribosomal Protein
P62847	RPS24	40S ribosomal protein S24	133	15.41	10.78	1.00	0.00	0.00	194.65	34.59	4.00	1.00	0.00	0.00	221.31	34.59	4.00	1.00	0.00	0.00	7.60	7.79	0.00	Ribosomal Protein
Q92552	MRPS27	28S ribosomal protein S27, mitochondrial	414	47.58	6.18	1.00	0.00	0.00	164.88	19.81	6.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	7.37	0.00	0.00	Ribosomal Protein
Q6DKI1	RPL7L1	60S ribosomal protein L7-like 1	246	28.64	10.51	1.00	0.00	0.00	147.10	17.07	3.00	1.00	0.00	0.00	138.28	12.20	3.00	1.00	0.00	0.00	7.20	7.11	0.00	Ribosomal Protein
P82933	MRPS9	28S ribosomal protein S9, mitochondrial	396	45.81	9.51	1.00	0.00	0.00	118.63	7.32	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.89	0.00	0.00	Ribosomal Protein
Q969Q0	RPL36AL	60S ribosomal protein L36a-like	106	12.46	10.65	1.00	0.00	0.00	118.43	36.79	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.89	0.00	0.00	Ribosomal Protein
P18077	RPL35A	60S ribosomal protein L35a	110	12.53	11.06	1.00	0.00	0.00	111.36	40.00	7.00	46.06	21.82	3.00	67.44	22.73	4.00	1.00	0.00	0.00	6.80	0.55	-5.53	Ribosomal Protein
Q9Y3B7	MRPL11	39S ribosomal protein L11, mitochondrial	192	20.67	9.91	1.00	0.00	0.00	102.12	26.56	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.67	0.00	0.00	Ribosomal Protein
P62979	RPS27A	Ubiquitin-40S ribosomal protein S27a	156	17.95	9.64	1.00	0.00	0.00	101.00	27.56	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.66	0.00	0.00	Ribosomal Protein
P82650	MRPS22	28S ribosomal protein S22, mitochondrial	360	41.25	7.90	1.00	0.00	0.00	99.66	12.50	3.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.64	0.00	0.00	Ribosomal Protein
Q9Y676	MRPS18B	28S ribosomal protein S18b, mitochondrial	258	29.38	9.38	1.00	0.00	0.00	84.20	12.40	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.40	0.00	0.00	Ribosomal Protein
P60866	RPS20	40S ribosomal protein S20	119	13.36	9.94	1.00	0.00	0.00	75.16	12.61	2.00	1.00	0.00	0.00	58.08	14.29	2.00	27.09	15.13	2.00	6.23	5.86	4.76	Ribosomal Protein
Q9Y3D9	MRPS23	28S ribosomal protein S23, mitochondrial	190	21.76	8.90	1.00	0.00	0.00	69.94	10.53	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	6.13	0.00	0.00	Ribosomal Protein
Q9Y3U8	RPL36	60S ribosomal protein L36	105	12.25	11.59	1.00	0.00	0.00	59.58	23.81	3.00	1.00	0.00	0.00	105.61	26.67	4.00	1.00	0.00	0.00	5.90	6.72	0.00	Ribosomal Protein
Q9Y2R5	MRPS17	28S ribosomal protein S17, mitochondrial	130	14.49	9.85	1.00	0.00	0.00	37.60	29.23	2.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	5.23	0.00	0.00	Ribosomal Protein
P61254	RPL26	60S ribosomal protein L26	145	17.25	10.55	41.89	10.34	2.00	331.66	56.55	13.00	39.22	11.03	2.00	236.81	46.90	11.00	1.00	0.00	0.00	2.99	2.59	-5.29	Ribosomal Protein
P27635	RPL10	60S ribosomal protein L10	214	24.59	10.08	67.07	22.90	3.00	505.51	56.54	13.00	1.00	0.00	0.00	297.35	38.32	7.00	1.00	0.00	0.00	2.91	8.22	0.00	Ribosomal Protein
P62424	RPL7A	60S ribosomal protein L7a	266	29.98	10.61	104.47	9.02	3.00	748.93	53.01	18.00	1.00	0.00	0.00	536.61	45.86	17.00	1.00	0.00	0.00	2.84	9.07	0.00	Ribosomal Protein
P39023	RPL3	60S ribosomal protein L3	403	46.08	10.18	156.23	18.86	6.00	1054.09	45.66	23.00	1.00	0.00	0.00	549.38	42.68	18.00	104.34	8.93	4.00	2.75	9.10	6.71	Ribosomal Protein
P62081	RPS7	40S ribosomal protein S7	194	22.11	10.10	56.34	21.65	2.00	349.61	56.70	11.00	71.01	14.95	3.00	189.21	46.91	6.00	78.31	17.53	2.00	2.63	1.41	0.14	Ribosomal Protein
P46779	RPL28	60S ribosomal protein L28	137	15.74	12.02	47.23	15.33	2.00	268.39	48.18	10.00	1.00	0.00	0.00	131.20	36.50	8.00	1.00	0.00	0.00	2.51	7.04	0.00	Ribosomal Protein
P05387	RPLP2	60S acidic ribosomal protein P2	115	11.66	4.54	72.48	39.13	2.00	345.35	85.22	6.00	1.00	0.00	0.00	264.77	73.91	7.00	1.00	0.00	0.00	2.25	8.05	0.00	Ribosomal Protein
P40429	RPL13A	60S ribosomal protein L13a	203	23.56	10.93	137.29	22.17	5.00	558.49	51.23	16.00	1.00	0.00	0.00	276.77	34.48	10.00	91.86	14.29	3.00	2.02	8.11	6.52	Ribosomal Protein
P42766	RPL35	60S ribosomal protein L35	123	14.54	11.05	68.17	18.70	2.00	268.59	29.27	4.00	1.00	0.00	0.00	126.30	18.70	2.00	1.00	0.00	0.00	1.98	6.98	0.00	Ribosomal Protein
P46776	RPL27A	60S ribosomal protein L27a	148	16.55	11.00	39.59	15.54	2.00	154.34	24.32	5.00	1.00	0.00	0.00	81.78	23.65	4.00	1.00	0.00	0.00	1.96	6.35	0.00	Ribosomal Protein
P26373	RPL13	60S ribosomal protein L13	211	24.25	11.65	129.88	22.27	5.00	497.96	44.55	13.00	1.00	0.00	0.00	236.26	37.44	9.00	98.92	8.06	2.00	1.94	7.88	6.63	Ribosomal Protein
P18621	RPL17	60S ribosomal protein L17	184	21.38	10.17	70.02	10.33	2.00	264.69	37.50	7.00	1.00	0.00	0.00	160.39	36.96	7.00	1.00	0.00	0.00	1.92	7.33	0.00	Ribosomal Protein
Q07020	RPL18	60S ribosomal protein L18	188	21.62	11.72	154.57	25.00	4.00	571.44	45.21	12.00	1.00	0.00	0.00	297.87	35.64	8.00	1.00	0.00	0.00	1.89	8.22	0.00	Ribosomal Protein
P61247	RPS3A	40S ribosomal protein S3a	264	29.93	9.73	127.93	26.14	6.00	470.32	43.94	11.00	53.13	9.47	3.00	423.54	43.56	10.00	68.87	9.09	3.00	1.88	2.99	0.37	Ribosomal Protein
P62917	RPL8	60S ribosomal protein L8	257	28.01	11.03	171.07	27.63	5.00	612.51	43.97	12.00	62.59	7.00	1.00	379.80	49.81	13.00	200.05	18.68	5.00	1.84	2.60	1.68	Ribosomal Protein
P36578	RPL4	60S ribosomal protein L4	427	47.67	11.06	294.14	30.21	12.00	979.59	50.59	28.00	1.00	0.00	0.00	800.16	49.88	24.00	1.00	0.00	0.00	1.74	9.64	0.00	Ribosomal Protein
P62906	RPL10A	60S ribosomal protein L10a	217	24.82	9.94	140.54	15.21	4.00	466.96	45.16	11.00	31.02	9.68	2.00	181.87	41.01	9.00	41.48	17.05	3.00	1.73	2.55	0.42	Ribosomal Protein
P62888	RPL30	60S ribosomal protein L30	115	12.78	9.63	114.23	24.35	2.00	373.06	64.35	8.00	103.45	40.87	3.00	354.52	64.35	8.00	112.17	40.87	3.00	1.71	1.78	0.12	Ribosomal Protein
P18124	RPL7	60S ribosomal protein L7	248	29.21	10.65	278.68	33.47	10.00	811.53	62.10	22.00	78.35	12.50	3.00	453.27	48.39	16.00	95.03	12.50	3.00	1.54	2.53	0.28	Ribosomal Protein
Q02878	RPL6	60S ribosomal protein L6	288	32.71	10.58	240.68	25.69	6.00	698.68	42.01	18.00	1.00	0.00	0.00	361.64	41.67	15.00	1.00	0.00	0.00	1.54	8.50	0.00	Ribosomal Protein
P46781	RPS9	40S ribosomal protein S9	194	22.58	10.65	144.59	23.20	6.00	419.32	51.55	16.00	55.77	8.25	2.00	329.29	47.94	15.00	89.55	12.37	3.00	1.54	2.56	0.68	Ribosomal Protein
P61313	RPL15	60S ribosomal protein L15	204	24.13	11.62	147.82	23.04	5.00	424.64	44.12	12.00	1.00	0.00	0.00	228.05	39.71	9.00	54.96	14.71	3.00	1.52	7.83	5.78	Ribosomal Protein
P32969	RPL9	60S ribosomal protein L9	192	21.85	9.95	96.34	22.40	3.00	273.66	40.10	6.00	1.00	0.00	0.00	171.03	30.21	5.00	1.00	0.00	0.00	1.51	7.42	0.00	Ribosomal Protein
P62750	RPL23A	60S ribosomal protein L23a	156	17.68	10.45	113.92	33.97	5.00	319.78	54.49	12.00	64.61	14.74	2.00	146.55	35.90	8.00	68.37	14.74	2.00	1.49	1.18	0.08	Ribosomal Protein
P15880	RPS2	40S ribosomal protein S2	293	31.30	10.24	136.07	28.67	7.00	358.15	40.27	10.00	1.00	0.00	0.00	243.15	44.03	11.00	1.00	0.00	0.00	1.40	7.93	0.00	Ribosomal Protein
P84098	RPL19	60S ribosomal protein L19	196	23.45	11.47	115.36	15.82	2.00	286.80	22.96	7.00	1.00	0.00	0.00	246.66	25.00	7.00	1.00	0.00	0.00	1.31	7.95	0.00	Ribosomal Protein
P49207	RPL34	60S ribosomal protein L34	117	13.28	11.47	66.39	16.24	3.00	162.11	30.77	6.00	1.00	0.00	0.00	44.90	14.53	2.00	1.00	0.00	0.00	1.29	5.49	0.00	Ribosomal Protein
P62753	RPS6	40S ribosomal protein S6	249	28.66	10.84	206.30	20.88	4.00	457.01	34.94	10.00	1.00	0.00	0.00	244.55	26.91	10.00	1.00	0.00	0.00	1.15	7.93	0.00	Ribosomal Protein
P46778	RPL21	60S ribosomal protein L21	160	18.55	10.49	76.88	24.38	3.00	167.27	29.38	5.00	57.95	13.13	2.00	168.35	35.00	5.00	74.09	13.13	2.00	1.12	1.54	0.35	Ribosomal Protein
P62899	RPL31	60S ribosomal protein L31	125	14.45	10.54	71.16	18.40	2.00	152.99	33.60	4.00	1.00	0.00	0.00	85.73	25.60	3.00	1.00	0.00	0.00	1.10	6.42	0.00	Ribosomal Protein
P62244	RPS15A	40S ribosomal protein S15a	130	14.83	10.13	158.31	50.77	6.00	339.10	53.85	7.00	92.58	35.38	5.00	311.98	67.69	10.00	98.65	35.38	5.00	1.10	1.75	0.09	Ribosomal Protein
Q02543	RPL18A	60S ribosomal protein L18a	176	20.75	10.71	203.34	35.80	7.00	434.98	43.75	11.00	1.00	0.00	0.00	239.23	42.05	9.00	52.18	9.66	3.00	1.10	7.90	5.71	Ribosomal Protein
P25398	RPS12	40S ribosomal protein S12	132	14.51	7.21	73.99	21.21	3.00	154.63	32.58	5.00	1.00	0.00	0.00	75.85	19.70	2.00	1.00	0.00	0.00	1.06	6.25	0.00	Ribosomal Protein
P62263	RPS14	40S ribosomal protein S14	151	16.26	10.05	101.82	16.56	3.00	205.01	66.23	7.00	58.64	15.89	2.00	129.34	56.95	5.00	1.00	0.00	0.00	1.01	1.14	-5.87	Ribosomal Protein
P83731	RPL24	60S ribosomal protein L24	157	17.77	11.25	129.33	19.11	3.00	249.11	31.85	5.00	1.00	0.00	0.00	113.93	18.47	4.00	1.00	0.00	0.00	0.95	6.83	0.00	Ribosomal Protein
P82930	MRPS34	28S ribosomal protein S34, mitochondrial	218	25.63	9.98	60.35	6.88	2.00	105.94	19.27	5.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.81	0.00	0.00	Ribosomal Protein
P62701	RPS4X	40S ribosomal protein S4, X isoform	263	29.58	10.15	376.46	43.35	13.00	653.13	52.09	15.00	127.01	19.01	5.00	559.66	51.71	15.00	256.46	37.64	10.00	0.79	2.14	1.01	Ribosomal Protein
P39019	RPS19	40S ribosomal protein S19	145	16.05	10.32	122.93	20.69	3.00	211.20	42.76	7.00	40.62	11.03	2.00	192.67	35.17	6.00	48.18	13.79	2.00	0.78	2.25	0.25	Ribosomal Protein
P62280	RPS11	40S ribosomal protein S11	158	18.42	10.30	152.41	34.81	5.00	258.24	50.63	8.00	69.07	25.95	5.00	225.26	62.03	12.00	99.24	26.58	5.00	0.76	1.71	0.52	Ribosomal Protein
P23396	RPS3	40S ribosomal protein S3	243	26.67	9.66	270.62	44.86	10.00	438.31	57.61	15.00	149.51	34.98	8.00	173.71	28.81	7.00	107.47	23.05	6.00	0.70	0.22	-0.48	Ribosomal Protein
P62854	RPS26	40S ribosomal protein S26	115	13.01	11.00	100.39	20.87	2.00	156.96	31.30	3.00	58.76	20.87	2.00	128.85	31.30	3.00	72.46	20.87	2.00	0.64	1.13	0.30	Ribosomal Protein
P62241	RPS8	40S ribosomal protein S8	208	24.19	10.32	298.62	37.02	6.00	457.07	51.44	9.00	159.86	22.60	4.00	428.92	49.04	8.00	219.55	29.33	6.00	0.61	1.42	0.46	Ribosomal Protein
P0CW22	RPS17L	40S ribosomal protein S17-like	135	15.54	9.85	161.35	39.26	5.00	237.49	49.63	6.00	1.00	0.00	0.00	210.83	39.26	5.00	1.00	0.00	0.00	0.56	7.72	0.00	Ribosomal Protein
P62249	RPS16	40S ribosomal protein S16	146	16.44	10.21	202.90	47.95	8.00	291.45	50.68	11.00	83.15	26.71	4.00	191.85	52.05	9.00	138.91	40.41	6.00	0.52	1.21	0.74	Ribosomal Protein
P62851	RPS25	40S ribosomal protein S25	125	13.73	10.11	143.51	29.60	5.00	186.67	28.80	5.00	71.88	15.20	2.00	124.00	29.60	5.00	1.00	0.00	0.00	0.38	0.79	-6.17	Ribosomal Protein
P62269	RPS18	40S ribosomal protein S18	152	17.71	10.99	337.06	57.89	12.00	412.89	57.89	13.00	266.81	43.42	8.00	357.03	57.89	12.00	342.81	54.61	12.00	0.29	0.42	0.36	Ribosomal Protein
P62266	RPS23	40S ribosomal protein S23	143	15.80	10.49	161.84	43.36	5.00	170.72	34.97	3.00	79.84	16.08	2.00	136.05	35.66	3.00	72.78	16.08	2.00	0.08	0.77	-0.13	Ribosomal Protein
P62891	RPL39	60S ribosomal protein L39	51	6.40	12.56	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	41.09	23.53	2.00	1.00	0.00	0.00	0.00	5.36	0.00	Ribosomal Protein

Accession	Gene Name	Description	# AAs	MW [kDa]	calc. pI	T7			T7-D8			F			F-D8			F-DR			Enrichment			Group
						Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	Score	Cover.	Uni. Pep.	T7-D8	F-D8	F-DR	
P83881	RPL36A	60S ribosomal protein L36a	106	12.43	10.58	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	108.48	35.85	5.00	1.00	0.00	0.00	0.00	6.76	0.00	Ribosomal Protein
P62277	RPS13	40S ribosomal protein S13	151	17.21	10.54	116.40	21.19	3.00	106.54	23.18	3.00	78.80	19.87	3.00	178.01	35.10	5.00	96.52	37.75	5.00	-0.13	1.18	0.29	Ribosomal Protein
P46782	RPS5	40S ribosomal protein S5	204	22.86	9.72	356.99	45.10	7.00	310.96	41.18	7.00	167.03	26.47	3.00	150.96	13.73	2.00	181.46	22.06	3.00	-0.20	-0.15	0.12	Ribosomal Protein
P11908	PRPS2	Ribose-phosphate pyrophosphokinase 2	318	34.75	6.61	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	49.77	7.55	2.00	0.00	0.00	5.64	Ribosomal Protein
P46783	RPS10	40S ribosomal protein S10	165	18.89	10.15	122.07	23.03	3.00	105.72	23.03	3.00	1.00	0.00	0.00	110.47	23.03	3.00	1.00	0.00	0.00	-0.21	6.79	0.00	Ribosomal Protein
P08865	RPSA	40S ribosomal protein SA	295	32.83	4.87	296.86	33.22	6.00	190.43	18.98	4.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	-0.64	0.00	0.00	Ribosomal Protein
P42677	RPS27	40S ribosomal protein S27	84	9.45	9.45	76.21	25.00	2.00	44.73	25.00	2.00	1.00	0.00	0.00	69.32	25.00	2.00	83.36	40.48	4.00	-0.77	6.12	6.38	Ribosomal Protein

Biochemical Society TRANSACTIONS

Biochemical Society Transactions

Cellular functions of the microprocessor

Sara Macias, Ross A. Cordiner and Javier F. Cáceres[1]

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, U.K.

Abstract

The microprocessor is a complex comprising the RNase III enzyme Drosha and the double-stranded RNA-binding protein DGCR8 (DiGeorge syndrome critical region 8 gene) that catalyses the nuclear step of miRNA (microRNA) biogenesis. DGCR8 recognizes the RNA substrate, whereas Drosha functions as an endonuclease. Recent global analyses of microprocessor and Dicer proteins have suggested novel functions for these components independent of their role in miRNA biogenesis. A HITS-CLIP (high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation) experiment designed to identify novel substrates of the microprocessor revealed that this complex binds and regulates a large variety of cellular RNAs. The microprocessor-mediated cleavage of several classes of RNAs not only regulates transcript levels, but also modulates alternative splicing events, independently of miRNA function. Importantly, DGCR8 can also associate with other nucleases, suggesting the existence of alternative DGCR8 complexes that may regulate the fate of a subset of cellular RNAs. The aim of the present review is to provide an overview of the diverse functional roles of the microprocessor.

Introduction

miRNAs (microRNAs) are small non-coding RNAs of approximately 22 nucleotides in length that negatively regulate the expression of complementary mRNAs. The mechanism of miRNA-mediated inhibition of gene expression operates at the level of mRNA stability and/or inhibition of translation [1]. miRNAs have diverse unique expression patterns and are involved in developmental and differentiation processes in animals, regulating a wide variety of cellular pathways [2]. It is becoming increasingly clear that miRNA biogenesis and function has to be tightly controlled and that aberrant miRNA expression can lead to the progression of human disease [3]. In the present article, we review the biogenesis pathway of miRNAs, with a particular emphasis on the recent discovery of non-canonical functions for the microprocessor.

miRNA biogenesis

Mammalian miRNAs are embedded in long pri-miRNAs (primary miRNAs), which are independent transcription units or can alternatively be located within introns of pre-mRNAs (precursor mRNAs). They are transcribed by RNA polymerase II and processing of pri-miRNAs by subsequent nuclear and cytoplasmic processing events gives rise to mature miRNAs. The nuclear step is catalysed by the microprocessor complex, comprising Drosha and its partner DGCR8 (DiGeorge syndrome critical region 8 gene) and results in the production of pre-miRNAs (precursor miRNAs). Subsequently, pre-miRNAs are exported from the nucleus by Exportin 5 and are further processed in the cytoplasm by the RNase III enzyme Dicer into mature miRNAs [4,5].

Biochemical studies of the microprocessor complex in human cells have shown that Drosha can form two different complexes. A smaller complex, comprising Drosha and its partner DGCR8, constitutes the minimal catalytically active complex *in vitro* that directs cleavage and production of pre-miRNAs [6,7]. A larger complex contains additional RNA-BPs (RNA-binding proteins), such as RNA helicases, hnRNPs (heterogeneous nuclear ribonucleoproteins) and other associated proteins that can regulate its activity [8]. DGCR8 is the microprocessor component that directly interacts with pri-miRNAs and functions as a molecular anchor that measures the distance from the dsRNA (double-stranded RNA)–ssRNA (single-stranded RNA) junction to direct the cleavage by the endonuclease Drosha 11 bp away from this junction [9,10]. In the cytoplasm, Dicer catalyses an additional processing step (dicing) that results in the production of approximately 22 nt miRNA duplexes. In humans, Dicer interacts with two closely related proteins, TRBP [TAR (transactivating response) RNA-binding protein; also known as TARBP2] and PACT (protein activator of the interferon-induced protein kinase; also known as PRKRA), which are not required for activity, but have been implicated in defining the cleavage site and facilitating the formation of the RISC (RNA-induced silencing complex) ([11], reviewed in [12]). One strand of this duplex is preferentially selected to bind to one of the Ago (Argonaute) proteins (Figure 1). Mammals have multiple Ago proteins, with Ago2 being the only one to have retained a slicer domain. The miRNA guides Ago to target mRNAs and can cause degradation or translation inhibition of the transcripts, depending on the extent of

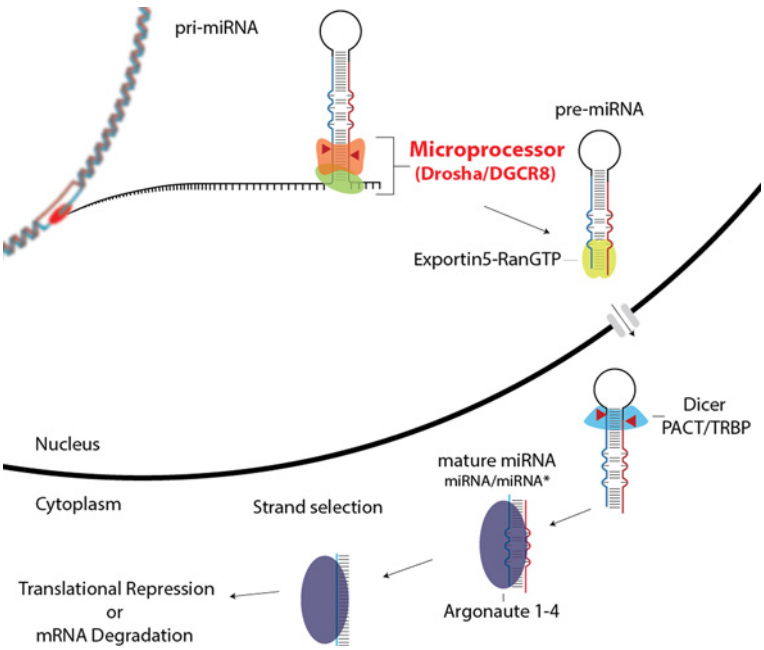
Key words: DGCR8, Dicer, Drosha, microprocessor, microRNA (miRNA), small RNA.

Abbreviations used: Ago, Argonaute; CTL, conserved terminal loop; DGCR8, DiGeorge syndrome critical region 8 gene; dsRNA, double-stranded RNA; ES, embryonic stem; HITS-CLIP, high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation; hnRNP, heterogeneous nuclear ribonucleoprotein; lincRNA, long intergenic non-coding RNA; miRNA, microRNA; pre-miRNA, precursor miRNA; pre-mRNA, precursor mRNA; pri-miRNA, primary miRNA transcript; RNA-BP, RNA-binding protein; sdRNA, snoRNA-derived RNA; siRNA, small interfering RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; ssRNA, single-stranded RNA; TAR, transactivating response; UTR, untranslated region.

[1]To whom correspondence should be addressed (email Javier.Caceres@igmm.ed.ac.uk).

Figure 1 | **miRNA biogenesis pathway**

RNA polymerase II transcribes long pri-miRNAs, which are processed co-transcriptionally. Nuclear cleavage by the microprocessor creates a 2-nt 3′ overhang that is recognized by Exportin 5-RanGTP, resulting in the export of the pre-miRNA out of the nucleus. The cytoplasmic processing step is carried out by Dicer, which generates a miRNA–miRNA* duplex. Only one strand will stably bind to Ago and then be used as a guide to target specific mRNAs for translational repression or degradation.



complementarity [1]. The canonical miRNA biogenesis pathway driven by the RNase III enzymes Drosha and Dicer has been challenged by the discovery of an unexpected variety of alternative mechanisms that generate functional miRNAs. These non-canonical pathways include, among others, the generation of mature miRNAs by the action of the spliceosome (mirtrons) (reviewed in [13]).

The production of mature miRNAs is tightly regulated at the post-transcriptional level at different steps throughout the biogenesis pathway. This can be achieved by proteins present in the large microprocessor complex and/or by auxiliary factors that directly bind to pre-miRNAs [14,15]. The most extensively studied example is the role of the RNA-BP Lin28 in the regulation of *let-7* miRNA family biogenesis, which is important for stem cell differentiation and normal development. Lin28 binds to the terminal loop of *let-7* family members and blocks its processing at both the Drosha and/or Dicer level [16]. Another interesting example is the activity of TGFβ (transforming growth factor β) and BMP (bone morphogenetic protein) signalling pathways that stimulate the Drosha-mediated processing of an individual miRNA *miR-21* [17]. In some cases, the same RNA-BP can display either stimulatory or inhibitory activities with different subsets of miRNAs. For instance hnRNP A1, a protein involved in many aspects of RNA processing, acts as an auxiliary factor for the Drosha-mediated processing of pri-*miR-18a*. It binds to the CTL (conserved terminal loop) of pri-*miR-18a* and rearranges its structure, creating a more favourable cleavage site for Drosha processing [18,19]. However, binding of hnRNP A1 to a CTL does not always lead to stimulation of miRNA processing. It was shown that hnRNP A1 is a negative regulator of *let-7a* in differentiated cells by antagonizing the positive role of KSRP (KH-type splicing regulatory protein) [20,21]. Altogether, these data suggest the existence of positive and/or negative auxiliary factors for the processing and production of specific miRNAs. Previous findings suggest that this type of regulation is extensive and in most cases involves recognition of conserved sequences in the terminal loop of pre-miRNAs (reviewed in [8]).

Non-canonical functions of the microprocessor

Functional studies of the miRNA processing machinery using mouse models showed that Drosha, DGCR8 and Dicer deficiencies resulted in all cases in early embryonic lethality, confirming a critical role for miRNAs in normal development [22–24]. MEFs (mouse embryonic fibroblasts) derived from these knockout mice showed some phenotypic differences as well as poor correlation between the populations of mRNAs affected by these two endonucleases. *Dicer*-null ES (embryonic stem) cells have normal morphology, but their proliferation potential is compromised [25,26]. *Dgcr8*-null ES cells cannot fully down-regulate pluripotency markers, which interferes with their ability to fully differentiate, suggesting that miRNAs have a crucial role in silencing ES cell self-renewal [23].

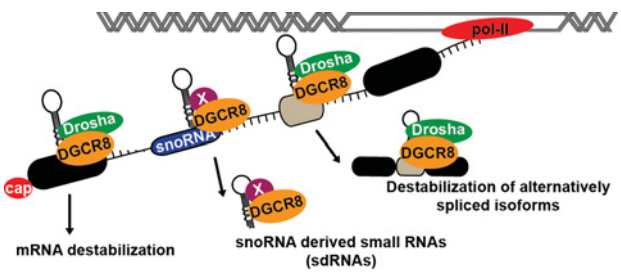
Transcriptional analyses of *Dicer*-null ES cells show that they are devoid of all canonical and most non-canonical miRNAs and siRNAs (small interfering RNAs), whereas *Dgcr8*-null ES cells are only lacking canonical miRNAs [23]. Thus these differences strongly suggested that Dicer has miRNA-independent roles in ES cell function. Furthermore, the fact that *Dgcr8*-null cells have less severe phenotypes than *Dicer*-null cells suggested an important role for microprocessor-independent, Dicer-dependent small RNAs. Indeed, profiling small RNAs for these two populations of cells identified DGCR8-independent, Dicer-dependent noncanonical miRNAs, which include mirtrons, endo-siRNAs and endo-shRNAs (small hairpin RNAs) [27].

Comparisons between *Dgcr8*/*Drosha* and *Dicer* knockout phenotypes were also pursued to elucidate their roles within differentiated cells and tissues, such as NK (natural killer) cells, regulatory thymocyte cells, epidermal and hair follicle epithelial cells [28–30]. These studies found that inactivation of Drosha/DGCR8 or Dicer gave rise to highly overlapping phenotypes, suggesting a common affected pathway. Nevertheless, in early T-cell progenitors, *Drosha* and *Dicer* conditional knockouts did not produce identical phenotypes, indicating non-redundant functions of Drosha and Dicer in miRNA processing during early T-cell development [24]. Transcriptomic changes that were unique to Drosha, but not Dicer, highlighted Drosha-dependent, Dicer-independent RNAs. Some of the mRNAs that were up-regulated in the absence of Drosha also contained a pri-miRNA-like structures which can be cleaved by the microprocessor, independently of miRNA production [24]. Indeed, Drosha had been shown to cleave pri-miRNA-like hairpins harboured within the 5′UTR (untranslated region) of the mRNA encoding the DGCR8 protein [31,32]. This negative regulatory feedback has also been reported in *Drosophila melanogaster*, indicating that this is a conserved regulatory pathway [33]. Comparison of the up-regulated mRNA populations in the absence of Drosha or Dicer identified another 25 transcripts that were proposed to be direct targets of the microprocessor. In addition, deep sequencing of Drosha cleavage products uncovered more mRNAs that could be regulated by endonucleolytic cleavage [34]. This regulation has been shown to be critical during mouse neurogenesis, where microprocessor-mediated destabilization of *Neurog2* mRNA facilitates neural stem cell maintenance [35]. The pre-mRNA encoding the KapB gene of KSHV (Karposi's sarcoma-associated herpesvirus) is also processed by the microprocessor as it contains a pri-miRNA located at the 3′UTR that also causes concomitant transcript destabilization [36]. It was reported recently that small RNAs that are recruited to the sites of DNA damage are Drosha- and Dicer-dependent; however, they are distinct from miRNAs [37]. These small RNAs, termed DDRNAs (DNA damage response RNAs), bind to double-strand breaks and recruit DDR (DNA damage response) proteins, leading to activation of the DNA damage response. Interestingly, defects in DNA damage repair have been seen before in miRNA processing knockouts [30]. Finally, microprocessor-mediated RNA cleavage can cause premature transcription termination of the TAR mRNA, encoded by HIV. In this case, Drosha-mediated endonucleolytic cleavage opens up the transcript for exonucleolytic degradation by Xrn2 and RRP6 [38]. Altogether, the evidence described above has strongly suggested an expanded role for non-canonical functions of the microprocessor and/or Dicer.

Direct RNA targets of the microprocessor complex

Computational scanning of the entire human genome searching for hairpin structures enabled the identification of approximately 11 million hairpins resembling validated pri-miRNA hairpins [39]. However, it is still not completely understood how the microprocessor complex selectively recognizes these structures. DGCR8 is the microprocessor component that binds and recognizes the dsRNA–ssRNA junction at the base of the pri-miRNA hairpin [10]; however, similar RNA secondary structures are found in many other cellular RNAs, suggesting that these could also be recognized and cleaved. The evidence described above pointed to a more extended role of the microprocessor in the regulation of a plethora of cellular RNAs. Recently, we took an alternative approach to identify direct RNA targets of the microprocessor complex, on the basis of sequencing the transcripts associated with DGCR8 by the HITS-CLIP (high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation) technique [40]. This technique relies on UV light irradiation of living cells to induce covalent cross-links between RNA and protein, allowing the use of very stringent washing conditions during the immunoprecipitation of the protein of interest to identify associated RNAs. It has been widely used to map *in vivo* RNA–protein interactions for several RNA-BPs involved in different aspects of RNA metabolism [41,42]. The study revealed that DGCR8 binds to many other types of structured RNAs other than miRNAs that harbour a predicted secondary structure resembling that of a pri-miRNA. These cellular targets comprised several hundred mRNAs, as well as snoRNAs (small nucleolar RNAs) and lincRNAs (long intergenic non-coding RNAs) (Figure 2) [40]. Direct binding was observed for more than 2000 mRNAs and depletion of the microprocessor resulted in an up-regulation of selected mRNAs, suggesting that these were direct RNA processing targets. Furthermore, *in vitro* cleavage by Drosha was confirmed for a handful of mRNA targets. Interestingly, DGCR8 was also shown to bind cassette exons, suggesting a role for this complex in the regulation of alternative splicing. Remarkably, in the absence of DGCR8 more than 300 alternatively spliced events were altered. It was found that the microprocessor cleaves and destabilizes mRNA isoforms harbouring DGCR8 binding sites in cassette exons, resulting in an altered ratio of alternatively spliced isoforms. The discovery that DGCR8 binding to cassette exons may act to influence the relative

Figure 2 | Non-canonical functions of the microprocessor
High-throughput sequencing of RNAs associated with DGCR8 (HITS-CLIP) revealed novel functions for the microprocessor. Drosha-mediated cleavage of exonic sequences at protein-coding mRNAs (black boxes) as well as lincRNAs induces destabilization of these transcripts (left). When binding occurs at an alternatively spliced (AS) cassette exons (grey box), this can influence the relative ratio between AS isoforms (right). DGCR8 also binds to snoRNAs (in blue) in association with another putative nuclease (X), other than Drosha, to generate sdRNAs or induce decay (centre).



abundance of alternatively spliced isoforms in a physiological manner is an exciting finding, but needs further investigation. In summary, this analysis revealed that the microprocessor binds and cleaves the mRNA of protein-coding genes, affecting mRNA abundance and also impacting alternative splicing regulation.

An alternative DGCR8 complex

DGCR8 targets include snoRNAs, short non-coding RNAs that modify rRNA, tRNA and snRNAs (small nuclear RNAs) by direct base pairing and recruitment of modifying enzymes, which guide 2′-O-methylation or pseudourydilation of the target RNA. Interestingly, snoRNAs are also a source of small RNAs, known as sdRNAs (snoRNA-derived RNAs) that associate with Ago proteins and act as functional miRNAs [43]. Whereas sdRNAs have been shown to be Dicer-dependent, but Drosha-independent in some cases [44], other reports have shown that sdRNAs are dependent on both DGCR8 and Dicer in several organisms, such as human, mouse, chicken, fruitfly, *Arabidopsis* and fission yeast [45]. Unexpectedly, DGCR8-mediated cleavage of snoRNAs is independent of Drosha, strongly suggesting that DGCR8 can associate with other nucleases to cleave subsets of target RNAs [40]. It may well be that snoRNAs are only one type of cellular RNA target recognized by this putative alternative DGCR8 complex. This opens an exciting avenue of new research aiming to identify additional nuclease(s) that associate with DGCR8 in alternative complexes and to identify cellular RNA targets of this complex(es).

Microprocessor functions and DiGeorge syndrome

The *DGCR8* gene is located in the 22q11.2 genomic region, which is deleted in DiGeorge syndrome patients [7,46]. This is the most common human genetic deletion syndrome that affects approximately one in 3000 live births. DiGeorge syndrome patients can present two different types of deletion, a 3.0 Mb or 1.5 Mb deletion, with the former being the most prevalent and spanning approximately 30 genes. In order to gain insight into this condition, an engineered mouse strain carrying a hemizygous chromosomal deficiency on chromosome 16 (that spans a segment syntenic to the human 1.5-Mb 22q11.2 microdeletion) was constructed. This $Dgcr8^{+/-}$ mutant mice [$Df(16)A^{+/-}$] displayed a minor alteration of miRNA production in the brain, but still exhibited behavioural and cognitive defects and cardiac abnormalities that resemble the human condition [47,48]. Further analysis of the $Df(16)A^{+/-}$ mice revealed abnormalities in the formation of neuronal dendrites and spines, as well as altered brain miRNAs. This phenotype was partially due to a drastic reduction in the levels of a single miRNA, *miR-185*, which resides within the 22q11.2 locus. Thus phenotypes observed in this mouse model could be attributed to the combined effect of *miR-185* and *Dgcr8* hemizygosity [49]. These studies highlighted an important role for miRNA biogenesis in this disease; however, it would be of great interest to determine whether altered non-canonical functions of the microprocessor could have any role in the origin and development of DiGeorge syndrome. The development of specific rescue experiments will unravel precisely which miRNAs are affected and whether the alternative functions of the microprocessor are necessary to restore brain function and revert the neuropsychiatric traits related to this deficiency.

Conclusions and future directions

In the present article, we reviewed evidence describing a more general role for the microprocessor that is not limited to the production of pre-miRNAs, but also directs the processing of a large number of cellular RNAs. The identification of non-canonical functions for the microprocessor complex highlights the necessity to reinterpret the phenotypes observed in DGCR8- and/or Drosha-deficient cells and animal models that were exclusively attributed to a defect in miRNA biogenesis. Evidence discussed in the present review also provides new insights into the complex role of the microprocessor in controlling gene expression by affecting the fate of several classes of RNAs.

Acknowledgements

We thank Elena Miñones-Moyano (Medical Research Council Human Genetics Unit) for comments on this review before submission.

Funding

This work was supported by core funding from the Medical Research Council and by the Wellcome Trust [grant number 095518/Z/11/Z].

References

1 Huntzinger, E. and Izaurralde, E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. **12**, 99–110
2 Bushati, N. and Cohen, S.M. (2007) microRNA functions. Annu. Rev. Cell Dev. Biol. **23**, 175–205
3 Mendell, J.T. and Olson, E.N. (2012) microRNAs in stress signaling and human disease. Cell **148**, 1172–1187
4 Winter, J., Jung, S., Keller, S., Gregory, R.I. and Diederichs, S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat. Cell Biol. **11**, 228–234
5 Krol, J., Loedige, I. and Filipowicz, W. (2010) The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. **11**, 597–610
6 Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004) The microprocessor complex mediates the genesis of microRNAs. Nature **432**, 235–240
7 Landthaler, M., Yalcin, A. and Tuschl, T. (2004) The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. Curr. Biol. **14**, 2162–2167
8 Siomi, H. and Siomi, M.C. (2010) Posttranscriptional regulation of microRNA biogenesis in animals. Mol. Cell **38**, 323–332
9 Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H. and Kim, V.N. (2004) The Drosha–DGCR8 complex in primary microRNA processing. Gene Dev. **18**, 3016–3027
10 Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T. and Kim, V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell **125**, 887–901
11 Fukunaga, R., Han, B.W., Hung, J.-H., Xu, J., Weng, Z. and Zamore, P.D. (2012) Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell **151**, 533–546
12 Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. **10**, 126–139
13 Yang, J.-S. and Lai, E.C. (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol. Cell **43**, 892–903
14 Newman, M.A. and Hammond, S.M. (2010) Emerging paradigms of regulated microRNA processing. Genes Dev. **24**, 1086–1092
15 Triboulet, R. and Gregory, R.I. (2010) Autoregulatory mechanisms controlling the microprocessor. Adv. Exp. Med. Biol. **700**, 56–66
16 Thornton, J.E. and Gregory, R.I. (2012) How does Lin28 let-7 control development and disease? Trends Cell Biol. **22**, 474–482
17 Davis, B.N., Hilyard, A.C., Lagna, G. and Hata, A. (2008) SMAD proteins control DROSHA-mediated microRNA maturation. Nature **454**, 56–61
18 Guil, S. and Cáceres, J.F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat. Struct. Mol. Biol. **14**, 591–596
19 Michlewski, G., Guil, S., Semple, C.A. and Cáceres, J.F. (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol. Cell **32**, 383–393
20 Michlewski, G. and Cáceres, J.F. (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nat. Struct. Mol. Biol. **17**, 1011–1018
21 Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R. and Rosenfeld, M.G. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature **459**, 1010–1014
22 Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V. and Hannon, G.J. (2003) Dicer is essential for mouse development. Nat. Genet. **35**, 215–217
23 Wang, Y., Medvid, R., Melton, C., Jaenisch, R. and Blelloch, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat. Genet. **39**, 380–385
24 Chong, M.M.W., Zhang, G., Cheloufi, S., Neubert, T.A., Hannon, G.J. and Littman, D.R. (2010) Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. **24**, 1951–1960
25 Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. **19**, 489–501
26 Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S. and Hannon, G.J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. Proc. Natl. Acad. Sci. U.S.A. **102**, 12135–12140
27 Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. and Blelloch, R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. **22**, 2773–2785
28 Chong, M.M.W., Rasmussen, J.P., Rudensky, A.Y., Rundensky, A.Y. and Littman, D.R. (2008) The RNAseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. J. Exp. Med. **205**, 2005–2017
29 Bezman, N.A., Cedars, E., Steiner, D.F., Blelloch, R., Hesslein, D.G.T. and Lanier, L.L. (2010) Distinct requirements of microRNAs in NK cell activation, survival, and function. J. Immunol. **185**, 3835–3846
30 Teta, M., Choi, Y.S., Okegbe, T., Wong, G., Tam, O.H., Chong, M.M.W., Seykora, J.T., Nagy, A., Littman, D.R., Andl, T. et al. (2012) Inducible deletion of epidermal Dicer and Drosha reveals multiple functions for miRNAs in postnatal skin. Development **139**, 1405–1416
31 Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.-K., Yeom, K.-H., Yang, W.-Y., Haussler, D., Blelloch, R. and Kim, V.N. (2009) Posttranscriptional crossregulation between Drosha and DGCR8. Cell **136**, 75–84
32 Triboulet, R., Chang, H.-M., Lapierre, R.J. and Gregory, R.I. (2009) Post-transcriptional control of DGCR8 expression by the microprocessor. RNA **15**, 1005–1011
33 Kadener, S., Rodriguez, J., Abruzzi, K.C., Khodor, Y.L., Sugino, K., Marr, M.T., Nelson, S. and Rosbash, M. (2009) Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. RNA **15**, 537–545
34 Karginov, F.V., Cheloufi, S., Chong, M.M.W., Stark, A., Smith, A.D. and Hannon, G.J. (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol. Cell **38**, 781–788
35 Knuckles, P., Vogt, M.A., Lugert, S., Milo, M., Chong, M.M.W., Hautbergue, G.M., Wilson, S.A., Littman, D.R. and Taylor, V. (2012) Drosha regulates neurogenesis by controlling Neurogenin 2 expression independent of microRNAs. Nat. Neurosci. **15**, 962–969
36 Lin, Y.-T. and Sullivan, C.S. (2011) Expanding the role of Drosha to the regulation of viral gene expression. Proc. Natl. Acad. Sci. U.S.A. **108**, 11229–11234
37 Francia, S., Michelini, F., Saxena, A., Tang, D., De Hoon, M., Anelli, V., Mione, M., Carninci, P. and d'Adda di Fagagna, F. (2012) Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature **488**, 231–235
38 Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O. et al. (2012) Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell **150**, 1147–1157
39 Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E. et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. **37**, 766–770
40 Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E. and Cáceres, J.F. (2012) DGCR8 HITS-CLIP reveals novel functions for the microprocessor. Nat. Struct. Mol. Biol. **19**, 760–766
41 Ule, J., Jensen, K., Mele, A. and Darnell, R.B. (2005) CLIP: a method for identifying protein–RNA interaction sites in living cells. Methods **37**, 376–386
42 Darnell, R.B. (2010) HITS-CLIP: panoramic views of protein–RNA regulation in living cells. Wiley Interdiscip. Rev.: RNA **1**, 266–286
43 Falaleeva, M. and Stamm, S. (2012) Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. BioEssays **35**, 46–54
44 Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N. and Meister, G. (2008) A human snoRNA with microRNA-like functions. Mol. Cell **32**, 519–28
45 Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P. and Mattick, J.S. (2009) Small RNAs derived from snoRNAs. RNA **15**, 1233–1240
46 Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. and Shimizu, N. (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. Biochem. Biophys. Res. Commun. **304**, 184–190

47 Stark, K.L., Xu, B., Bagchi, A., Lai, W.-S., Liu, H., Hsu, R., Wan, X., Pavlidis, P., Mills, A.A., Karayiorgou, M. et al. (2008) Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat. Genet. **40**, 751–760
48 Fénelon, K., Mukai, J., Xu, B., Hsu, P.-K., Drew, L.J., Karayiorgou, M., Fischbach, G.D., Macdermott, A.B. and Gogos, J.A. (2011) Deficiency of *Dgcr8*, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. Proc. Natl. Acad. Sci. U.S.A. **108**, 4447–4452
49 Xu, B., Hsu, P.-K., Stark, K.L., Karayiorgou, M. and Gogos, J.A. (2013) Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. Cell **152**, 262–275

Received 1 February 2013
doi:10.1042/BST20130011